UNED

Técnicas e instrumentos para la recogida de información

Juan Antonio Gil Pascual

COLECCIÓN GRADO

Técnicas e instrumentos para la recogida de información

JUAN ANTONIO GIL PASCUAL

UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA

TÉCNICAS E INSTRUMENTOS PARA LA RECOGIDA DE INFORMACIÓN
6302205GR01A01

Quedan rigurosamente prohibidas, sin la autorización escrita de los titulares del Copyright, bajo las sanciones establecidas en las leyes, la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la reprografía y el tratamiento informático, y la distribución de ejemplares de ella mediante alquiler o préstamos públicos.

© *Universidad Nacional de Educación a Distancia*
Madrid 2011

Librería UNED: c/ Bravo Murillo, 38 - 28015 Madrid
Tels.: 91 398 75 60 / 73 73
e-mail: libreria@adm.uned.es

© *Juan Antonio Gil Pascual*

Todas nuestras publicaciones han sido sometidas a un sistema de evaluación antes de ser editadas.

ISBN: 978-84-362-6250-6
Depósito legal: 33.376-2011

Primera edición: agosto de 2011

Impreso en España - Printed in Spain

ÍNDICE

Capítulo 1. **La medida en Educación** 13
1. Introducción 15
2. La medida en Educación 16
   2.1. La medición y la técnicas estadísticas 21
3. Técnicas de recogida de datos 23
4. La informática y las técnicas de recogida de información 23
Anexo: breve manual de R 25
5. Entorno de trabajo y carga de módulos de R 25
   5.1. Introducción 25
   5.2. Paquetes 25
   5.3. Documentación 26
   5.4. Ayuda y comentarios 27
6. El R como calculadora 27
7. Estructuras de datos en R: vectores, factores, matrices y arrays, listas y data frames 29
   7.1. Objetos en R 29
   7.2. Vectores 33
   7.3. Matrices 37
8. Lectura y escritura de datos 38
   8.1. Lectura de datos 38
      8.1.1. De un fichero 38
      8.1.2. La función scan() 39
      8.1.3. Leer datos de librerías 39
      8.1.4. Acceder a datos de una biblioteca 39
   8.2. Exportar datos 40
   8.3. Guardar y leer datos 40
9. Graficos 41
   9.1. Introducción 41
10. Estructura de programación y funciones 47
   10.1. Definición de función 47
   10.2. Control de ejecución 47

11. La exploración de datos: estadísticos y funciones de distribución de probabilidad 48
11.1. La exploración de los datos 48
11.2. Distribuciones de probabilidad 49
11.3. Funciones para estadísticas simples 50
12. Diversos contrastres estadísticos 51
Bibliografía 53

Capítulo 2. **Características técnicas de los instrumentos de medida** 55
1. Características técnicas de los instrumentos de medida 57
2. Teoría clásica de los test (TCT) 59
2.1. Dificultad 59
2.2. Discriminación 60
2.3. Análisis de distractores 63
2.4. Índice de fiabilidad 63
2.5. Validez 69
2.6. Dimensionalidad 72
3. Teoría de respuesta al ítem 73
3.1. Supuestos 74
3.2. Modelos 74
3.3. Curva característica del test 79
3.4. Funciones de información 81
3.4.1. Ponderación óptima de los ítems 84
4. Aplicaciones de la TRI 85
5. Un ejemplo de cálculo de las características técnicas de un test 87
Bibliografía 104

Capítulo 3. **La observación** 105
1. La observación 107
2. El observador 108
3. Tipos de observación 109
4. Unidades de observación 111
5. El proceso de observar 113
6. Técnicas para el registro de los datos de observación 115
6.1. Lista de control 116
6.2. Registro anecdótico 116
6.3. Sistemas de rasgos 117

6.4. Escalas de estimación o apreciación ... 117
6.5. Sistemas de categorías ... 118
7. Unidades de medida ... 120
8. Análisis de validez y fiabilidad ... 120
8.1. Fiabilidad como acuerdo entre observadores (fiabilidad inter-observadores) ... 121
8.2. Fiabilidad entre observaciones ... 122
9. Análisis de datos ... 122
10. Un ejemplo de observación en la práctica docente ... 124
Bibliografía ... 126

Capítulo 4. **Test y pruebas objetivas** ... 129
1. Introducción ... 131
2. Test ... 131
3. Pruebas objetivas ... 132
4. Construcción de las pruebas ... 132
4.1. Objetivos de los test o pruebas ... 133
4.2. Especificación de la prueba ... 134
4.3. Confección de la prueba ... 138
4.3.1. Confección de los ítems ... 138
4.3.2. Confección de la prueba ... 139
5. Aplicación de la prueba ... 142
6. Puntuación de la prueba ... 143
6.1. Puntuaciones de los ítems ... 143
6.2. Puntuación de la prueba ... 144
7. Interpretación de las puntuaciones de la prueba ... 144
8. Clasificación de los test ... 147
Bibliografía ... 149

Capítulo 5. **Técnicas de encuestas** ... 151
1. Introducción ... 153
2. Tipo de encuestas ... 153
2.1. La encuesta personal ... 155
2.2. La encuesta telefónica ... 156
2.3. La encuesta total ... 158
2.4. Las encuestas por ordenador y on-line ... 161
3. Criterios en la selección del tipo de encuesta ... 164
4. Proceso de una encuesta ... 166

4.1. Diseño de cuestionarios 166
4.1.1. Reglas para la formulación de preguntas 167
4.1.2. Tipos de preguntas 170
4.1.3. Estructura del cuestionario 173
4.1.4. La codificación en el cuestionario 174
4.1.5. Estudio de las características técnicas 175
4.1.6. Comprobación y prueba del cuestionario. Pretest 176
4.2. Diseños muestrales 178
4.2.1. Selección de la muestra 179
4.3. El trabajo de control 179
4.3.1. Introducción 179
4.3.2. Recomendaciones en el proceso de la entrevista 179
4.3.3. Armonía entre el entrevistador y el encuestado 179
4.3.4. Formulación de las preguntas 180
4.3.5. Registro de respuesta 181
4.3.6. Falseamiento de las encuestas 182
4.3.7. Control del trabajo de campo 182
4.4. El análisis estadístico 183
4.4.1. Tratamiento y análisis de la información 183
4.5. El informe de resultados 184
4.6. Ayudas informáticas para el tratamiento de encuestas 187
4.6.1. Programas 187
4.7. Conclusiones 188
5. Un ejemplo de encuesta 190
Bibliografía 197

Capítulo 6. **La entrevista** 199
1. Introducción 201
2. Características de la entrevista 202
3. Fases de una entrevista 203
4. Tipos de entrevista 204
4.1. La entrevista estructurada 204
4.2. Entrevista no estructurada 207
4.3. Entrevista de grupo o grupal 211
5. Análisis de datos 211
5.1. Tamaño de la muestra 212
6. Un ejemplo de entrevista 212
Bibliografía 216

Capítulo 7. **Las escalas** ... 219
1. Las escalas ... 221
2. Escalas de actitudes ... 224
   2.1. Escalas de Likert ... 225
   2.2. Escalas de Thurstone ... 229
   2.3. Escalas de Guttman ... 233
3. Escalas de producción escolar ... 236
4. Diferencial semántico ... 236

Bibliografía ... 241

Capítulo 8. **Pruebas criteriales** ... 243
1. Pruebas criteriales ... 245
2. Tipos de pruebas criteriales ... 246
3. Elaboración de pruebas criteriales de clase ... 247
   3.1. Especificación del dominio educativo ... 247
   3.2. Análisis de los ítems ... 249
   3.3. Determinación de estándares y puntos de corte ... 252
   3.4. Estudio de las características técnicas de la prueba ... 252
4. Un ejemplo de prueba criterial ... 253

Bibliografía ... 254

Capítulo 9. **Técnicas grupales de recogida de información** ... 255
1. Introducción ... 257
2. Tipos ... 257
3. Grupos focales ... 257
4. Grupo nominal ... 259
5. Entrevistas en profundidad ... 260
6. El Brainstorming ... 261
   6.1. Requisitos de una sesión de brainstorming o tormenta de ideas ... 262
   6.2. Brainstorming paso a paso ... 262
7. El método Delphi ... 264
   7.1. Fases ... 265
8. Técnicas sociométricas ... 267
   8.1. El sociograma ... 268
   8.2. Escala de distancia social ... 273
   8.3. Los inventarios de personalidad ... 276

Bibliografía ... 277

Capítulo 10. **Análisis de documentos** 279
1. Análisis de documentos 281
2. Análisis de contenido 282
3. Tipos de documentos 284
4. Procedimiento a seguir en el análisis de contenido 284
5. Tipologías de análisis de contenido 287
   5.1. Análisis de contenido temático 287
   5.2. Análisis de contenido semántico 289
   5.3. Análisis de contenido de redes 290
6. Ejemplo de análisis de contenido 292

Bibliografía 302

Capítulo 1
LA MEDIDA EN EDUCACIÓN

1. Introducción
2. La medida en Educación
   2.1. La medición y la técnicas estadísticas
3. Técnicas de recogida de datos
4. La informática y las técnicas de recogida de información

Anexo: breve manual de R

5. Entorno de trabajo y carga de módulos de R
   5.1. Introducción
   5.2. Paquetes
   5.3. Documentación
   5.4. Ayuda y comentarios
6. El R como calculadora
7. Estructuras de datos en R: vectores, factores, matrices y arrays, listas y data frames
   7.1. Objetos en R
   7.2. Vectores
   7.3. Matrices
8. Lectura y escritura de datos
   8.1. Lectura de datos
   8.2. Exportar datos
   8.3. Guardar y leer datos
9. Graficos
   9.1. Introducción
10. Estructura de programación y funciones
    10.1. Definición de función
    10.2. Control de ejecución
11. La exploración de datos: estadísticos y funciones de distribuciónde probabilidad
    11.1. La exploración de los datos
    11.2. Distribuciones de probabilidad
    11.3. Funciones para estadísticas simple
12. Diversos contrastres estadísticos

Objetivos

- Comprender la utilidad de la medición en educación
- Comprender el concepto de medida y los tipos de escala de medida
- Valorar la importancia de la informática y la estadística como apoyo en la construccón de instrumentos de medida
- Repasar la utilización de R en el tratamiento de datos

1. INTRODUCCIÓN

La ciencia se apoya en instrumentos capaces de cuantificar fenómenos estableciendo relaciones funcionales obtenidas a través de modelizaciones de la realidad.

Para realizar este proceso, el investigador utiliza la medida mediante la que es capaz de representar propiedades por medio de números. Ciertas propiedades han sido estudiadas desde hace siglos: los procesos de medida del tiempo, la longitud o el peso, son de todos conocidos, pero otras características, no menos importantes, como la inteligencia, la atención, el interés, etc. han sido investigadas recientemente.

Para algunos, la medida de ciertos atributos del acto educativo es imposible de cuantificar. Sin entrar en una polémica estéril, resulta incuestionable la diferencia en cantidad de información y grado de exactitud existente, por ejemplo, entre la medida del peso de un alumno y la medida de su capacidad intelectual.

Sin embargo, es interesante observar cómo las ciencias han progresado a partir del momento en que se ha dispuesto de buenos instrumentos de medida. Conviene tener presente la regla de Galileo: "medir lo medible e intentar hacer medible lo que aún no lo es", sobre todo si aceptamos el principio del matemático y filósofo H. Poincaré, según el cual "lo que no se mide no puede ser objeto de ciencia". Sin embargo, la necesidad de medir no significa que todo sea medible o que todo se pueda medir, se necesita conocer las variables a medir y someterlas a prueba, aventurar si entre sus características se encuentra la posibilidad de ser medida.

Siguiendo las premisas expresadas, en este capítulo se presentan los rudimentos de medida en el ámbito educativo y un breve anexo del programa estadístico R como apoyo a la medición.

2. LA MEDIDA EN EDUCACIÓN

Los procesos de medida física vienen siendo abordados desde comienzos del siglo XX con una axiomática capaz de justificar, por ejemplo, la medida de masas. Los procedimientos que intervienen en la medida del peso de un estudiante son sencillos: para calcularlo, se establece una relación de orden — "tan pesado como"— de manera que si se denomina p la escala de medida y p(x) el número que asigna a cada estudiante su peso, se cumplirá que: a) $p(x) \geq p(y)$ si y sólo si $x \geq y$; b)denotando por $x*y$ la combinación de dos elementos, se cumple $p(x*y)=p(x)+p(y)$, aditividad de peso.

La asignación de numerales a entes físicos ha sido muy estudiada, sin embargo, el proceso de asignación a elementos de ciencias sociales no. Su medida se tiene que basar en un sistema de preferencias, de utilidades. Supongamos que queremos medir la capacidad de transmisión de información de unos profesores a través de las opiniones de sus alumnos. Si éstos clasifican a sus docentes según sus opiniones respecto a la característica señalada, podemos definir también una ordenación de la siguiente manera: denominando con v(x) la valoración realizada por los alumnos del profesor x; la utilidad que se le asigne debe reflejar tanto la ***ordenación*** de los profesores, $v(x) \geq v(y)$ si y sólo si $x \geq y$, como la de las ***diferencias*** entre profesores, $[v(x)-v(y)] \geq [v(z)-v(w)]$ si y sólo si $(x,y) \geq (z,w)$.

Aunque no podemos medir directamente la inteligencia, la motivación, el estrés, etc., nos valemos de indicadores del comportamiento a medir. En este sentido un ***indicador*** es un instrumento que pone de manifiesto alguna característica no observable directamente (Bisquerra, 1989 p.89). En palabras de Bunge (1976, p.802) "nunca se mide una cosa: lo medible es exclusivamente alguna propiedad de la cosa".

En el mundo de la medición en ciencias sociales se está trabajando mucho en los modelos de ***rasgo latente*** que, mediante el establecimiento de una relación matemática entre las habilidades de un sujeto y los ítems de un cuestionario, test o prueba, permite inferir de forma probabilística el éxito de un individuo en una tarea, de forma que la medida no se ve influida por el instrumento utilizado, uno de los mayores inconvenientes de la medición en ciencias sociales.

Otro concepto a destacar en la medida en Educación es el de ***evaluación***. No es una medida simple, sino que lleva inmerso un proceso de emitir juicio

de valor sobre las mediciones realizadas. La evaluación puede recaer en aspectos próximos al rendimiento académico y en otros relacionados con la investigación evaluativa.

Todo proceso de asignación de numerales a entidades está inmerso en dos posibles casuísticas: una ***relación de orden,*** donde se establece una dominancia de un elemento sobre otro, y una ***relación de proximidad*** en que se valora la distancia que separa a las entidades.

Por otra parte, los dos problemas más importantes que se manejan en la teoría de medición son: a) el problema de ***representación***, que trata de demostrar la existencia de una representación numérica para la valoración, y b) el problema de ***unicidad***, que aborda la caracterización de las representaciones numéricas existentes. De estos dos problemas, es el segundo el que más afecta a la investigación educativa, de ahí que sea abordado a continuación.

En la teoría de la medición en Ciencias Sociales hay dos tendencias, a nuestro juicio, no excluyentes: la ***teoría clásica*** de la ***medición*** representada por la figura de Stevens y el ***enfoque axiomático*** de la ***medición***.

Para Stevens la medición es *"la asignación de números a objetos o eventos de acuerdo con leyes"* y en consecuencia distingue diferentes tipos de asignación. Así el tipo o nivel de medida se clasifica en función del tipo de transformación que deja invariante la forma de la escala. Es decir, para Stevens las propiedades de las observaciones empíricas están ligadas al tipo de escala a utilizar.

Según el criterio de Stevens los distintos niveles de medida son: nominal, ordinal, intervalo y razón de relación (ratio).

Medidas de tipo nominal

En la tipología de Stevens, este tipo de medidas es el de grado inferior, puesto que los valores de la variable no tienen ninguna relación de nivel o distancia entre ellos. Cada valor es una categoría distinta, y sólo sirve de etiqueta o nombre (de aquí la palabra "nominal") para la categoría. ***No asume orden ni distancia entre las categorías creadas***. Por ejemplo, el sexo o la nacionalidad, son variables nominales.

En nuestra opinión cuando asignamos numerales al sexo o la nacionalidad, propiamente no estamos creando una escala, porque no podemos comparar con otro elemento patrón o referencia. Realmente estamos numerando o clasificando elementos de conjuntos.

Medidas de tipo ordinal

Cuando es posible establecer un ***orden*** *o* ***rango*** dentro de las categorías de acuerdo con un criterio, estamos dentro de las medidas ordinales.

Este tipo de medidas debe abordar los problemas de la clasificación de las categorías y de la medida de las distancias entre ellas.

Medidas de intervalo

En las medidas de intervalo, en orden ascendente de posibilidades de operación, ***la distancia entre las categorías está definida en unidades iguales y fijas***. Por ejemplo, un termómetro mide la temperatura en grados y cada medición pertenece a una escala. Así la diferencia entre 30 y 31 grados centígrados es la misma que entre 80 y 81 grados farenheit. En las medidas de intervalo lo que se mide es la diferencia entre ellos, es decir, en una escala de intervalo se puede pasar de una escala (x) a otra (y) por una transformación lineal.

$$y=ax+b \ (a\neq 0)$$

Como se ve, el origen cero de la escala no siempre es el mismo (cero relativo).

Si consideramos cuatro valores x_1, x_2, x_3, x_4 y los correspondientes y_1, y_2, y_3, y_4 se verifica:

$$\frac{y_1 - y_2}{y_3 - y_4} = \frac{(ax_1 + b) - (ax_2 + b)}{(ax_3 + b) - (ax_4 + b)} = \frac{x_1 - x_2}{x_3 - x_4}$$

es decir, la razón de intervalos es invariante.

En investigación social es difícil encontrar verdaderas medidas de intervalo.

Medidas de Ratio (Razón)

Las medidas de nivel ratio tienen todas las propiedades de las medidas de intervalos y además una propiedad adicional: el nivel cero absoluto es inherente al esquema de esta medida. Así, cuando medimos distancias físicas, si usamos centímetros o metro, el cero es una medida naturalmente definida: es la ausencia de distancia entre dos objetos. Esta propiedad de fijar y de dar el punto cero mediante la comparación de ratio es una buena medida de distancia. Por ejemplo, se puede decir con exactitud que un hombre de 1,80 metros de estatura es dos veces la talla de un niño que mide 90 centímetros, es decir, se puede establecer una ***proporcionalidad***, una razón.

Como vemos se puede pasar de una escala (x) a otra (y) por una transformación:

$$y=cx \quad (c\neq 0)$$

Hay por tanto, en estas escalas un cero absoluto y se pasa de una medida a otra multiplicando por una constante. Si tenemos dos valores x_1, x_2 y los correspondientes y_1, y_2 , se cumplirá:

$$\frac{y_1}{y_2}=\frac{cx_1}{cx_2}=\frac{x_1}{x_2}$$

lo que justifica el nombre dado a estas escalas.

Además, las medidas de ratio satisfacen todas las propiedades del sistema de números reales.

En investigación educativa las variables que manejamos llegan, en el mejor de los casos, a ser de intervalo. Sólo determinadas variables físicas de los sujetos (pero sin relevancia o con poca relevancia desde el aspecto educativo, como el peso, la altura, etc.) alcanzan la medida de nivel de razón.

La teoría de Stevens ha sido durante muchos años del pasado siglo fuertemente apoyada por psicólogos y en general en investigación social. A pesar de lo cual, la definición de Stevens es errónea porque confunde dos clases diferentes de prácticas: **medida** (en el sentido clásico) y la **codificación** numérica. Por medida entendemos un compromiso filosófico de la cantidad y el número como parte de la utilería del universo, mientras que la codificación numérica es simplemente un recurso natural para la representación simbólica de los hechos.

Por otra parte, el **enfoque axiomático** considera a la **medición** como la construcción de un modelo de alguna propiedad del mundo. Como toda modelización, se quiere establecer una correspondencia entre un sistema relacional empírico (la realidad) y un sistema relacional formal (modelo). Así, dado el sistema relacional a=(A,R), con A un conjunto de objetos y R una relación definida en ese conjunto, se dice que está representado por el conjunto b=(B,S), si existe una correspondencia f de A en B tal que para todo x,y de A se cumple:

$$xRy \Rightarrow f(x)Sf(y)$$

Si se cumple que *a* representa a *b* y *b* representa a *a,* se dice que los sistemas son isomorfos.

Desde el enfoque axiomático, la esencia de la representación numérica es la asignación de números a objetos, de tal manera que las relaciones observadas entre los objetos queden reflejadas por las correspondientes relaciones entre los números asignados.

Desde el punto de vista axiomático existen tantas escalas como conjunto de relaciones aritméticas se puedan encontrar. Además, la incertidumbre sobre el tipo de escala se puede resolver mediante la comparación con otra escala conocida.

Las últimas evoluciones de este enfoque axiomático nos lleva a un énfasis sobre las variables cuantitativas, medir y medida. Las variables cuantitativas tienen dos características definitorias: el **orden** y su **estructura aditiva**. Se puede decir que medir es un procedimiento para identificar valores de variables cuantitativas a través de relaciones numéricas con otros valores.

Si U es la unidad de medida, y mido X puedo decir que X=cU, o lo que es lo mismo, X/U=c. Así los valores que no conozco los infiero del valor de la unidad y se pueden establecer relaciones como X=cU e Y=kU, entonces X/Y=c/k. Las medidas relativas a la misma unidad forman una escala de medida, y las medidas en la misma escala tienen en cuenta además relaciones numéricas para ser obtenidas.

La medida requiere el desarrollo de procedimientos mediante los cuales X y U puedan compararse y obtener su razón. Dichos procedimientos son los métodos de medida, y el desarrollo de tales métodos se denomina cuantificación.

Naturalmente antes de cuantificar hay que comprobar que la variable es cuantitativa. Si este paso no se ha realizado el resto del proceso de medición es puramente una especulación.

Con el objetivo de conseguir esta comprobación debemos imponer a cualquier medida la independencia entre la variable de interés y el instrumento utilizado para medirla. Además, en las variables sociales, un elemento primordial es la probabilidad al ser la conducta humana no determinista.

El modelo de teoría de respuesta al ítem o rasgo latente de Rasch, contempla estos dos aspectos: la separabilidad y la convergencia de la probabilidad. Dicho modelo examina la posibilidad de someter a comprobación la naturaleza cuantitativa de las variables.

La ecuación básica de un proceso de medida planteado por Rasch se fundamenta en la idea de que la respuesta a un ítem sólo depende de la capacidad del sujeto (θ) y la dificultad intrínseca del ítem (b). Y plantea el siguiente modelo logístico:

$$P_i(\theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}} \text{ donde}$$

$P_i(\theta)$: es la probabilidad de acertar el ítem i a un determinado nivel θ

θ: Valores de la variable medida

b_i: Índice de dificultad del ítem i

Si los datos se ajustan al modelo de Rasch la variable tiene naturaleza cuantitativa y si los datos no se ajustan a dicho modelo, son los datos los que deben cambiarse, no el modelo (Gaviria, J.L. 2000, p. 222). Aunque con una visión menos estricta que la del autor, en nuestra opinión pensamos que la distribución de los datos marca, de alguna forma, la naturaleza de la variable.

2.1. La medición y las técnicas estadísticas

Los principios de medición del enfoque Stevens conducen a una concepción acerca de la relación entre medición y estadística. Según este planteamiento la escala de medida determina por completo qué tipo de análisis estadístico es el más apropiado. Así, plantea la siguiente ***tabla 1.1***, que luego

se verá reflejada en manuales de pruebas estadísticas, como el clásico de Siegel (1970).

Tabla 1.1. Técnicas estadísticas apropiadas según la escala de medida (Muñiz 2000, p. 292)

Escala	Medida de posición	Dispersión	Asociación o correlación	Prueba de significación
Nominal	Moda	Información, H	Información transmitida, T coeficiente de contingencia	Ji-cuadrado
Ordinal	Mediana	Percentiles	Correlación ordinal	Prueba de los signos
Intervalo	Mediana arirtmética	Desviación típica	Correlación producto-momento	Prueba T Prueba F
Razón	Media geométrica Media aritmética	Variación porcentaje		

En contra de la postura de Stevens, se han alzado otras voces de estadísticos como Anderson (Muñiz, 2000, p. 292): "*la validez de las inferencias estadísticas no puede depender del tipo (o no tipo) de escala de medida utilizada*".

No obstante, aunque el significado empírico de los números no afecte a la validez de las manipulaciones realizadas sobre ellos, sí afecta a las conclusiones extraídas de los resultados de esas manipulaciones. Pero esta afirmación es sólo parte del proceso, porque se pueden estar utilizando, por ejemplo, *medidas de intervalo* como la relajación muscular, tiempo empleado en comunicación con los compañeros, etc. como índice de satisfacción escolar y querer sacar inferencias de un hecho educativo *ordinal*: la satisfacción.

El tema, como vemos, es controvertido y seguro que seguirá despertando polémica. No obstante, en algunas cuestiones fijamos nuestra posición:

1) Stevens confundía la teoría de la medición con la teoría estadística y subordinaba los procedimientos estadísticos al tipo de escala de medición utilizada.

2) Los procedimientos estadísticos no requieren propiedades específicas de la escala.

3) Los procedimientos estadísticos sólo están limitados por aspectos matemáticos derivados del modelo asumido por la técnica estadística utilizada. Por ejemplo, la prueba ANOVA (análisis de la varianza) asume el modelo lineal (normalidad, independencia, homocedasticidad, etc.).

4) Las conclusiones en el uso de un procedimiento estadístico vienen delimitadas por el contexto donde se aplica

3. TÉCNICAS DE RECOGIDA DE DATOS

El concepto de ***técnicas*** de recogida de información engloba todos los medios técnicos que se utilizan para registrar las observaciones o facilitar el tratamiento. Dentro de la expresión "medios técnicos" están inmersos, por una parte, los ***instrumentos*** —objetos con entidad independiente y externa— y los ***recursos*** —medios utilizados para obtener y registrar la información—. Por ejemplo, utilizando como técnica el cuestionario, podemos manejar como instrumento un cuestionario cerrado, empleando como recurso el ordenador.

Respecto a la clasificación de las técnicas de recogida de datos, no existe acuerdo entre distintos autores: Pérez Juste (1986), García Jiménez (1994) y Angulo (1990), entre otros; no obstante, recogiendo sus concordancias, básicamente se pueden dividir en seis grandes grupos: ***la observación, entrevista, cuestionario, pruebas objetivas y tests, técnicas grupales y sociométricas y análisis de documentos***.

El investigador siempre debe elegir aquel instrumento que más se ajuste al diseño de investigación planteado, considerando, a su vez, que debe conservar las características imprescindibles de un instrumento de medida: validez, fiabilidad y grado de operatividad de las variables.

4. LA INFORMÁTICA Y LAS TÉCNICAS DE RECOGIDA DE INFORMACIÓN

En algunos casos, como medio de recogida de información, se utilizan medios electrónicos. Así tenemos el caso de las grabaciones de audio reco-

gidas en las entrevistas o tomas en video realizadas mediante cámaras. Esta forma de recoger información presenta la ventaja añadida de la escasa intervención del observador y la perdurabilidad del momento captado, que permite una posible repetición y valoración.

Independientemente de los medios para recoger información, la informática juega un papel importante como soporte de datos. Las bases de datos, los editores o las hojas de cálculo son elementos normalmente utilizados para estos menesteres. Estos medios informáticos están frecuentemente integrados en paquetes estadísticos, que abordan, como una fase más del tratamiento, la recogida de datos; de esta forma, el filtrado de la información, la codificación y modificación, son etapas de esta fase de entrada de datos.

Otro punto donde interviene la informática es en los llamados test computarizados y adaptativos, derivados de la teoría de respuesta al ítem. Son test adaptados a los niveles particulares de aptitud de los sujetos (Martínez Arias, 1995, p. 319).

Para el cálculo de las características técnicas de los instrumentos de medida, para análisis estadísticos de los datos, para gráficos, cálculo del tamaño muestral, generación de valores aleatorios, etc., se necesita la ayuda de programas estadísticos. El programa R, software libre de análisis estadístico de datos, es una herramienta de suma utilidad para realizar los cálculos, en alguno de los cuales resulta de obligado uso (teoría de respuesta al ítem). El lector puede consultar un breve manual en anexo y en la página web del programa referencia a distintos manuales.

ANEXO: BREVE MANUAL DE R

5. ENTORNO DE TRABAJO Y CARGA DE MÓDULOS DE R

5.1. Introducción

R es un software estadístico-matemático orientado al proceso de datos, con un entorno gráfico de trabajo. Se pueden destacar las siguientes componentes:

- Un sistema de manejo y almacenamiento de datos eficiente.
- Funciones matemáticas para el proceso de vectores y matrices.
- Una gran cantidad de funciones estadísticas integradas en el sistema.
- Una amplia variedad de librerías especializadas para estadística y análisis de datos.
- Funciones de proceso gráfico orientadas al análisis de datos.
- Un lenguaje de programación completo basado en el lenguaje S.

El lenguaje de programación R, es un dialecto del lenguaje S, desarrollado en AT&T por Becker, Chambers y Wilks. En general, cualquier manual de aprendizaje de S o SPlus servirá para aprender a programar en R.

R se obtiene de forma gratuita en http://www.r-project.org/. En dicha dirección se pueden encontrar distintos materiales de apoyo, no obstante el lector que desee bajarse una versión "portable", es decir, que sólo necesita descomprimir un fichero .zip y ejecutar el fichero R_Portable.exe, puede utilizar la página web del autor. En esta versión portable tiene además, el R-Comander para manejar los procedimientos más habituales de R sin necesidad de aprender los comandos.

5.2. Paquetes

R consta de un sistema base y de paquetes adicionales que permiten una mayor funcionalidad.

Uno de los paquetes más interesante para realizar los trabajos estadísticos mediante una interface gráfica, y por tanto sin necesidad de conocer comandos, es: Rcmdr. Si se baja la versión portable de R ya está instalado dicho paquete.

Cuando sea necesario utilizar un paquete que no está dentro del sistema base será necesario cargarlo. Si el paquete no ha sido instalado, en el interface gráfico de R-Console(Gui) será necesario realizar la siguiente secuencia: Paquetes→Instalar paquetes(s)→Seleccionar CRAN mirror→Seleccionar el paquete deseado.

Una vez instalado el paquete siempre que se quiera usar será necesario, por ejemplo, dar la orden:

>library(ctest)[1]

donde ctest es el nombre de una librería que se desea usar especialmente dedicada a pruebas de contraste de hipótesis.

5.3. Documentación

Los manuales de R incluidos en todas las instalaciones son:

- *An introduction to R*. (De lectura obligatoria)
- Writing R extensions
- R data import/export
- The R language definition
- R installation and administration

Además de los múltiples documentos en inglés, recomendamos consultar http://cran.r-project.org/other-docs.html#nenglish donde se pueden encontrar distintos materiales de apoyo en idioma distinto al inglés.

[1] El símbolo > indica el puntero de la consola (Gui), naturalmente no debe escribirse dicho símbolo al comienzo del comando.

5.4. Ayuda y comentarios

Existen diversos procedimientos de ayuda:

Para encontrar todos los comandos que contengan un determinado término:

>apropos (“normal”)

Para saber más de la sintaxis de un comando:

>help (“if”)

Para poner un comentario aclaratorio en alguna instrucción se utiliza el símbolo # que hace invisible al programa el texto colocado a la derecha del símbolo.

```
>a<-c(3,4,8,16) ;b<-c(7,3,6,12)
>a+b # presenta en consola el vector resultado de la suma de a+b
```

6. EL R COMO CALCULADORA

R puede ser utilizado como una calculadora de modo interactivo. Por ser los comienzos, en este apartado nos limitaremos a mostrar algunos ejemplos sencillos.

Ejemplo de comando completo en R :

>2+2	Escribimos 2+2 y damos enter
[1] 4	Sale el resultado. [1] indica que es el primer (y único resultado) de nuestra orden.
>	**Al terminar el comando, el sistema vuelve a presentar su indicador. Si se hace a través de R-Comander no presenta ningún indicador en la ventana de sintaxis.**

Otros ejemplos:

> 2*5 [1] 10	Multiplicación de dos números
> 5/2 [1] 2.5	División real de dos números

```
> 5%/%2          División entera: se devuelve la parte entera solamente
[1] 2
> 5%%2           Módulo: resto de dividir un número por otro
[1] 1
> 11%%3          Otro ejemplo de la operación módulo
[1] 2
> 5^2            Potenciación
[1] 25
> 5^2.3          Potenciación, ex“““““nente real
[1] 40.51641
> exp(1)         El número e
[1] 2.718282
> exp(3)         El número e³
[1] 20.08554
> sqrt(2)        La raíz cuadrada de 2
[1] 1.414214
> log(3)         El logaritmo neperiano de 3
[1] 1.098612
> log(3,10)      El logaritmo de 3 en base 10
[1] 0.4771213
> abs(-3.4)      Valor absoluto de un número
[1] 3.4
> pi             El número pi
[1] 3.141593
```

Para averiguar un poquito más sobre la organización de las funciones matemáticas en R, podemos probar las siguientes órdenes:

```
> help(“Math”)
```

Que da un conjunto de funciones que podemos agrupar de la siguiente manera:

Arith: “+”, “-”, “*”, “^”, “%%”, “%/%”, “/”
Compare: “==”, “>”, “<”, “!=”, “<=”, “>=”
Logic: “&”, “|”, but not “!” since that has only one argument.
Ops: “Arith”, “Compare”, “Logic”
Math: “log”, “sqrt”, “log10”, “cumprod”, “abs”, “acos”, “acosh”, “asin”, “asinh”, “atan”, “atanh”, “ceiling”, “cos”, “cosh”, “cumsum”, “exp”, “floor”, “gamma”, “lgamma”, “sin”, “sinh”, “tan”, “tanh”, “trunc”

Math2: “round”, ”signif”
Summary: “max”, “min”, “range”, “prod”, “sum”, “any”, “all”
Complex : “Arg”, “Conj”, “Im”, “Mod”, “Re”
>help(“Special”)

Con las siguientes funciones:

beta(a, b)
lbeta(a, b)
gamma(x)
lgamma(x)
psigamma(x, deriv = 0)
digamma(x)
trigamma(x)
choose(n, k)
lchoose(n, k)
factorial(x)
lfactorial(x)

donde a, b, x, n son vectores numéricos y k, deriv son vectores enteros.

7. ESTRUCTURAS DE DATOS EN R: VECTORES, FACTORES, MATRICES Y ARRAYS, LISTAS Y DATA FRAMES

7.1. Objetos en R

Como R es un lenguaje que maneja objetos, casi todo en R es un objeto, esto significa que las variables, datos, funciones, resultados, etc., se guardan en la memoria activa del ordenador en forma de *objetos* con un *nombre* específico. El usuario puede modificar o manipular estos objetos con *operadores* (aritméticos, lógicos, y comparativos) y *funciones* (que a su vez son objetos).

Tipos de objetos:

- Objetos del lenguaje: llamadas, expresiones, nombres
- Funciones: constan de lista de argumentos, código, entorno
- Sin objeto: NULL

Manipulación de los objetos de la memoria:

ls() lista los objetos de la memoria. Por ejemplo:

```
> name <- "Juan"; n1 <- 5; n2 <- 50; p <- 0.5
> ls()
[1] "p"    "n1"    "n2"    "name"
```

pat si se quiere listar los objetos que contengan un carácter en particular. Por ejemplo:

```
> ls(pat="n")
[1] "n1"    "n2"    "name"
```

La función ***ls.str()*** muestra algunos detalles de los objetos en memoria:

```
> ls.str()
n1 : num 5
n2 : num 50
name : chr "Juan"
```

Para borrar objetos en memoria, utilizamos la función ***rm()***:

```
> rm(n1)
```

Objetos para los datos:

vector: colección ordenada de elementos del mismo tipo.

```
> x <- c(1, 2, 3); y <- c("a", "b", "c")
> z <- c(TRUE, TRUE, FALSE)
```

array: generalización multidimensional del vector. Elementos del mismo tipo.

```
> x <- array(1:3,c(2,4))
```

data frame: como el array, pero con columnas de diferentes tipos. Es el objeto más habitual para los datos experimentales.

```
>resul<-data.frame(ID=c("alum1","alum2","alum3","alum4"),
peso=c(54,53,48,58),altura=c(1.62,1.65,1.68,1.7),CI=c(98,86,89,90))
```

Como resultado de salida es:

```
> resul
    ID   peso altura CI
```

```
1 alum1  54  1.62 98
2 alum2  53  1.65 86
3 alum3  48  1.68 89
4 alum4  58  1.70 90
```

Si interesa referirse exclusivamente a los datos de peso se expresará de la siguiente forma:

```
> resul$peso
[1] 54 53 48 58
```

factor: tipo de vector para datos categóricos.

```
> x <- factor(c(1, 2, 2, 1, 1, 2, 1, 2, 1))
```

Se puede discretizar datos usando las funciones ***split*** y ***cut*** , veamos:

```
> n<-10;n1<-10
> g<-factor(letters[1:5],label="a")
> x <- rnorm(n * n1) + sqrt(as.numeric(g))
> xg <- split(x, g)
> xg
```

```
$a1
[1] -0.5544007 1.7491799 1.2400475 -0.4611344 0.9476911 1.0546162
2.2115396 0.7194146 0.9430771 -0.6691185 2.2864430 -0.4671794

[13] 1.9664175 -0.4061556 0.1318291 1.2546443 1.0511046 1.4302350
1.4453412 -0.6818947

$a2
[1] 0.5932663 1.2285783 1.4956471 -0.3039916 -0.6588399 0.4301719 -
0.4255259 0.3838169 2.8414031 0.7472850 3.4598941 3.6798054

[13] 1.3957887 3.2903877 0.6314057 0.6212782 1.7107844 1.0903122
1.8117973 0.9504550

$a3
[1] 1.639137642 -0.002936988 1.045994480 1.389934483 1.829202854
1.569667022 1.281325420 3.123278536 0.490092245 0.859599867
[11] 1.436474666 2.358285698 1.457578449 0.757540936 2.187309621
0.641944190 0.879695564 1.416748460 0.637300954 1.968718943
```

```
$a4
[1] 1.1977327 2.3174723 2.4618665 2.2352206 0.7690333 3.7358991
1.2880352 1.0780257 3.0511156 1.5542524 1.2816942 3.1158970
2.8098809
[14] 1.4647825 3.1957905 2.7698156 1.9034218 1.5698684 3.5791493
2.9078323

$a5
[1] 3.1900371 3.9673368 3.1072440 1.7157778 2.4530234 1.9962088
1.9643648 1.0355729 0.5030147 2.2691015 3.3046722 1.4398403
2.7860069
[14] 1.7853105 2.4721768 2.7556695 1.3202772 1.4886730 1.4092787
2.0698601
```

```
> v1<-c(7,8,10,1,2,5)
> cut1<-cut(v1,3);summary(cut1)
(0.991,4]   (4,7]   (7,10]
        2       2        2
```

list: vector generalizado. Cada lista está formada por componentes, que pueden ser otras listas, y cada componente puede ser de distinto tipo. Son contenedores generales de datos. Muy flexibles, pero sin estructura. Muchas funciones devuelven una lista o conjunto de resultados de distinta longitud y distinto tipo.

```
una.lista<-
c(un.vector=1:5,una.palabra="gracias",una.matriz=matrix(round(rn
orm(20),3),ncol=5),lista2=c(a=3,b=factor(c("a","b"))))
>una.lista
```

Como resultado de salida es:

```
un.vector1  un.vector2  un.vector3  un.vector4  un.vector5 una.palabra
una.matriz1
   "1"     "2"     "3"     "4"     "5"  "gracias"   "-0.671"
una.matriz2 una.matriz3 una.matriz4 una.matriz5 una.matriz6
una.matriz7 una.matriz8
  "0.846"   "0.77"   "1.085"   "0.704"   "0.557"   "1.124"   "-1.462"
```

```
una.matriz9 una.matriz10 una.matriz11 una.matriz12 una.matriz13
una.matriz14 una.matriz15
  "-2.015"   "-1.2"   "0.346"   "1.164"   "-1.33"   "-0.858"   "0.672"
una.matriz16 una.matriz17 una.matriz18 una.matriz19 una.matriz20
lista2.a  lista2.b1
  "0.879"   "0.82"   "1.771"   "1.29"   "0.317"    "3"    "1"
 lista2.b2
    "2"
```

7.2. Vectores

— Generación de secuencias

> x <- c(1, 2, 3, 4, 5)
> x <- 1:10; y <- -5:3
> x <- seq(from = 2, to = 18, by = 2)
> x <- seq(from = 2, to = 18, length = 30) #donde length indica el número de puntos de la secuencia
> y <- seq(along = x) #donde along indica la secuencia de puntos del argumento x
> z2 <- c(1:5, 7:10, seq(from=-7,to=5,by=2))

Cuya salida es:

> z2
[1] 1 2 3 4 5 7 8 9 10 -7 -5 -3 -1 1 3 5
> rep(1, 5)
> x <- 1:3; rep(x, 2)
> y <- rep(5, 3); rep(x, y)
> rep(1:3, rep(5, 3))
> rep(x, x)
> rep(x, length = 8)
> gl(3, 5) # como rep(1:3, rep(5, 3))
> gl(4, 1, length = 20) # Tener en cuenta que gl genera factores
> gl(3, 4, label = c("Infantil", "Primaria", "Secundaria"))
> expand.grid(edad = c(10, 18, 25), genero = c("Hombre", "Mujer"), loc = 1:3)

Cuya salida es:

```
  edad genero loc
1   10 Hombre  1
2   18 Hombre  1
3   25 Hombre  1
4   10  Mujer  1
5   18  Mujer  1
6   25  Mujer  1
7   10 Hombre  2
8   18 Hombre  2
9   25 Hombre  2
10  10  Mujer  2
11  18  Mujer  2
12  25  Mujer  2
13  10 Hombre  3
14  18 Hombre  3
15  25 Hombre  3
16  10  Mujer  3
17  18  Mujer  3
18  25  Mujer  3
```

Se puede combinar:

z5 <- c(1:5, rep(8, 3))

— Generación de secuencias aleatorias

```
> sample(5)
> sample(5, 3)
> x <- 1:15
> sample(x)
> sample(x, replace = TRUE)
> sample(x, length<- 2* length(x), replace = TRUE)
> probs <- x/sum(x)
> sample(x, prob = probs)
```

Números aleatorios rDistribución(n,parámetros)

```
> rnorm(10) # rnorm(10, mean = 0,sd = 1)
> runif(8,min=2,max=10)
```

— Selección de elementos de un vector

```
> x <- 1:5; x[1]; x[3]; x[c(1,3)]
> x[x > 3]
> x > 3 # Da como salida valores lógicos
> y <- x > 3
> x[y]
> x[-c(1, 4)]; y <- c(1, 2, 5); x[y]
> names(x) <- c("a", "b", "c", "d", "uva")
> x[c("b", "uva")]
```

— Valores perdidos

Todo objeto tiene dos atributos *intrínsecos*: *tipo* y *longitud*. El tipo se refiere a la clase básica de los elementos en el objeto; existen cuatro tipos principales: numérico, carácter, complejo, y lógico (FALSE [Falso] o TRUE [Verdadero]). Existen otros tipos, pero no representan datos como tal (por ejemplo funciones o expresiones). La longitud es simplemente el número de elementos en el objeto. Para ver el tipo y la longitud de un objeto se pueden usar las funciones ***mode*** y ***length***, respectivamente:

```
> x<-24
> mode(x)
[1] "numeric"
> length(x)
[1] 1
```

NA cuando un dato no está disponible (del inglés "not available").

```
> v <- c(1,6,9,NA)
> is.na(v); which(is.na(v))
> w <- v[!is.na(v)] # sin los valores perdidos
```

Sustituir NA por, p. ej., 0:

```
> v[is.na(v)] <- 0
```

En R se representa correctamente valores numéricos no-finitos como ±∞ con Inf y -Inf, o valores que no son numéricos con NaN (del inglés "not a number").

```
> 5/0; -5/0; 0/0
> is.infinite(-5/0); is.nan(0/0); is.na(5/0)
```

Con algunas funciones

> xna <- c(1, 2, 3, NA, 4); mean(xna)
> mean(xna, na.rm = TRUE) # calcula la media de xna despreciando los valores NA

— Eliminar todos los NA:

> xnb <- matrix(c(1,2,NA,3,NA,4), nrow = 3)
> xnb
> x.no.na <- na.omit(xnb)

— Ordenación de vectores

> xa <- c(5, 1, 8, 2)
> order(xa) # orden en que aparecen los datos
> sort(xa) # ordenar los datos en orden creciente
> rank(xa) # calcula los rangos
> xa[order(xa)] # vector con los elementos ordenados de forma creciente
> xb <- c(1, 2, 2, 3, 3, 4); rank(xb) # observar cómo se calcula el rango en caso de empate
> min(xa)
> which.min(xa); which(xa == min(xa)) # lugar que ocupa un elemento dado
> y <- c(1, 1, 2, 2); order(y, x)
order y sort admiten decreasing = TRUE.

— Vectores de caracteres

> codigos <- paste(c(“X”, “Y”), 1:2, sep = “”)
> codigos <- paste(c(“X”, “Y”), 1:2, sep = “.”)
> juntar <- paste(c(“un”, “día”, “feliz”, collapse =””))
> columna.a <- LETTERS[1:5]; columna.b <- 20:25
> juntar <- paste(columna.a, columna.b, sep =””)
> substr(“abcdef”,2,3)
> x <- paste(LETTERS[1:5], collapse=””)
> substr(x, 3, 5) <- c(“pa”) # sustituye las letras que ocupan el lugar 3º y 4º de x por las letras pa, quedando por tanto como resultado de x: ABpaE

7.3. Matrices

Una *matrix* es un *array* con dos dimensiones. Se puede utilizar de forma parecida pero *matrix* es más cómodo de emplear.

```
> x1 <- array(7, dim = c(3,2))
> x1c<- array('pepe',dim=c(3,2))
> x2 <- matrix(1:40, nrow = 5) # orden de formación por columnas
> x3 <- matrix(1:40, nrow = 5, byrow = TRUE) #orden de formación por filas
> x4 <- 1:14; dim(x4) <- c(7, 2)
```

Operaciones con matrices:

A %*% B	producto de matrices
t(A)	transpuesta de la matriz A
solve(A,b)	solución del sistema de ecuaciones Ax=b.
solve(A)	inversa de la matriz A
svd(A)	descomposición en valores singulares
qr(A)	descomposición QR
la.qr <- qr(A) ; names (la.qr)	uno de los atributos de la.qr es el rango y se localiza asi:

```
> print(c('el rango de la matriz es',la.qr$rank),quote=F)
```

eigen(A)	valores y vectores propios
diag(B)	matriz diagonal
sum(diag(B))	traza de una matriz
diag(3)	definir una matriz identidad
all(B == t(B))	comprobar si una matriz es simétrica
det(B)	determinante de una matriz
A %o% B == outer(A,B)	producto exterior de dos vectores o matrices
rbind(pa,pb,pc)	concatena por filas las matrices pa, pb y pc
cbind(pa,pb)	concatena por columnas las matrices pa y pb
apply(ma,1,mean)	calcula la media de las filas de la matriz ma
apply(ma,2,mean)	calcula la media de las columnas de la matriz ma
tapply(x,y,mean)	calcula las medias de la variable x en las categorías de la variable y. Por ejemplo

```
> x<-c(18,29,22,19,16,17,25,27,30,23,28)
```

```
> cat<-c(rep("A",5),rep("B",6))
> media.x<-tapply(x,cat,mean)
> media.x
   A    B
20.8 25.0
```

La siguiente **tabla resume** los tipos de objetos y los datos que representan.

Objeto	Tipo	¿Varios tipos posibles en el mismo objeto?
vector	numérico, carácter, complejo, lógico	No
factor	numérico o carácter	No
array	numérico, carácter, complejo, lógico	No
matriz	numérico, carácter, complejo, lógico	No
data.frame	numérico, carácter, complejo, lógico	Si
ts	numérico, carácter, complejo, lógico	Si
lista	numérico, carácter, complejo, lógico, expresión, función	Si

8. LECTURA Y ESCRITURA DE DATOS

8.1. Lectura de datos

8.1.1. De un fichero

read.table lee un fichero con datos separados por espacios en blanco, tabuladores o saltos de línea.

```
read.table(file, header = FALSE, sep = "", quote = "\"", dec = ".",row.names, col.names, as.is = FALSE, na.strings = "NA",
colClasses = NA, nrows = -1,skip = 0, check.names = TRUE, fill = !blank.lines.skip,strip.white = FALSE, blank.lines.skip = TRUE,
comment.char = "#")
```

Para poder leer una hoja de datos directamente, el archivo externo debe reunir las condiciones adecuadas. La forma más sencilla es:

- La primera línea del archivo debe contener el nombre de cada variable de la hoja de datos.
- En cada una de las siguientes líneas, el primer elemento es la etiqueta de la fila, y a continuación deben aparecer los valores de cada variable.

```
> tabla.datos <- read.table("c:/R/mis.datos", header = TRUE, sep = "",comment.char = "")
```

8.1.2. La función scan()

Lee datos de un fichero o de la consola de trabajo. Su formato es:

```
scan(file = "", what = double(0), nmax = -1, n = -1, sep = "",
   quote = if(identical(sep, "\n")) "" else "\"", dec = ".",
   skip = 0, nlines = 0, na.strings = "NA",
   flush = FALSE, fill = FALSE, strip.white = FALSE,
   quiet = FALSE, blank.lines.skip = TRUE, multi.line = TRUE,
   comment.char = "", allowEscapes = FALSE, encoding = "unknown")
```

8.1.3. Leer datos de librerías

Con la función data se puede leer los ficheros de datos de las librerías disponibles en el programa.

```
data() # muestra todos los ficheros
> data(co2)
> data(co2, package = "datasets") # es equivalente
> co2 # presenta en consola los datos
```

8.1.4. Acceder a datos de una biblioteca

Para acceder a los datos incluidos en una biblioteca, basta utilizar el argumento ***package*** en la función ***data***. Por ejemplo,

```
> data(package="nls")
> data(Puromycin, package="nls")
```

Si la biblioteca ya ha sido conectada mediante la función ***library***, sus datos habrán sido incluidos automáticamente en la trayectoria de búsqueda y no será necesario incluir el argumento ***package***. Así,

```
> library(nls)
> data()
> data(Puromycin)
```

8.2. Exportar datos

Para exportar datos lo más sencillo es exportar una matriz (es necesario transponer la matriz).

```
> write(t(x), file = "c:/R/data.txt", ncolumns = n, append = FALSE)
```

Pero para exportar un ***data.frame*** es mejor

```
> write.table(my.data.frame,file = "mi.output.txt", sep = "",row.names = FALSE, col.names = TRUE)
```

Para escribir un fichero CSV importable desde Excel

```
> write.table(x, file = "foo.csv", sep = ",", col.names = NA)
```

8.3. Guardar y leer datos

Resulta muy importante poder guardar datos, funciones, etc., para ser usados en otras sesiones de R. Esos datos así guardados pueden compartirse con otros usuarios e incluso utilizarse en distintos sistemas operativos.

```
> x <- runif(30)
> y <- list(a = 1, b = TRUE, c = "salida")
> save(x, y, file = "xy.RData")
```

Los leeremos con

```
> load("xy.RData")
```

Podemos salvar todos los objetos con

> save.image() # salvado como ".RData"
> save.image(file = "un.nombre.RData")

9. GRÁFICOS

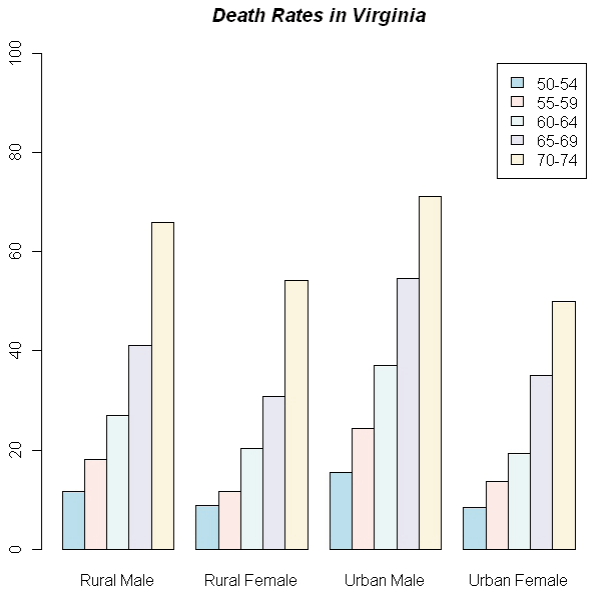
9.1. Introducción

Se puede ver diferentes posibilidades gráficas dando:

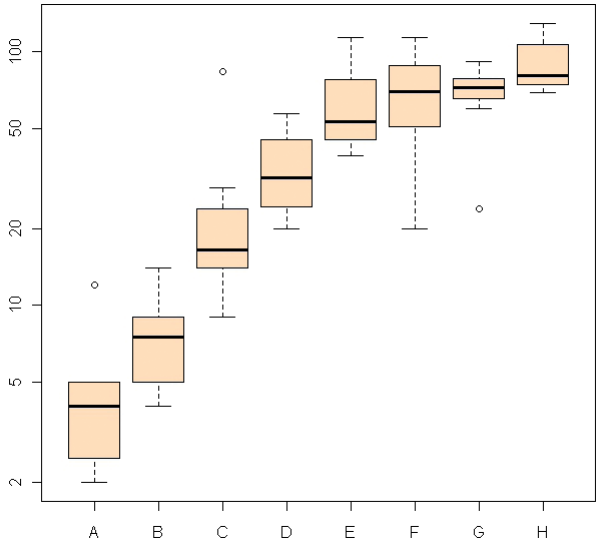
> demo(graphics)

Gráficos alto nivel:

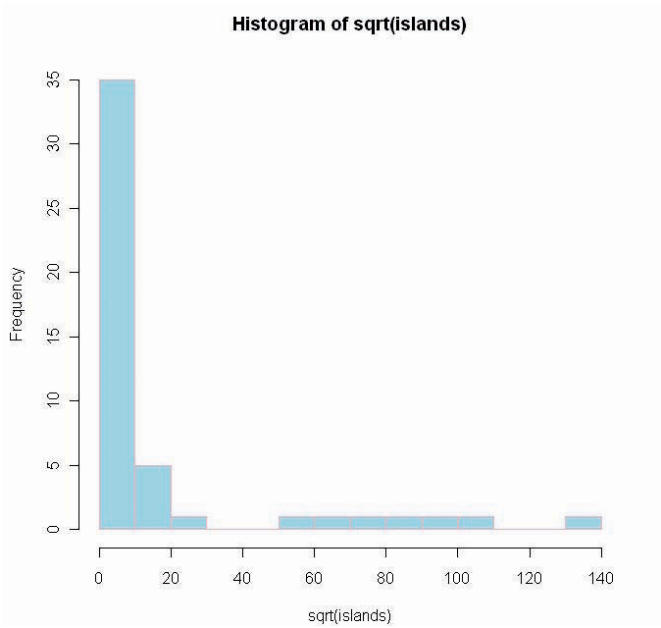
barplot(VADeaths, beside = TRUE, col = c("lightblue", "mistyrose", "lightcyan", "lavender", "cornsilk"), # **diagrama de barras**
legend = rownames(VADeaths), ylim = c(0, 100)) # leyenda del diagrama de barras
title(main = "Death Rates in Virginia", font.main = 4) # titulo del diagrama de barras



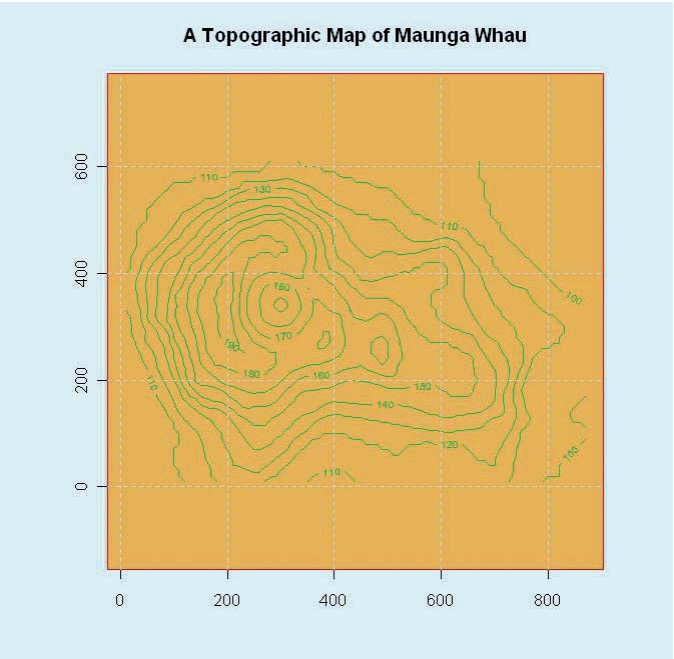
boxplot(decrease ~ treatment, data = OrchardSprays, log = “y”, col = “bisque”) # **diagrama de cajas**



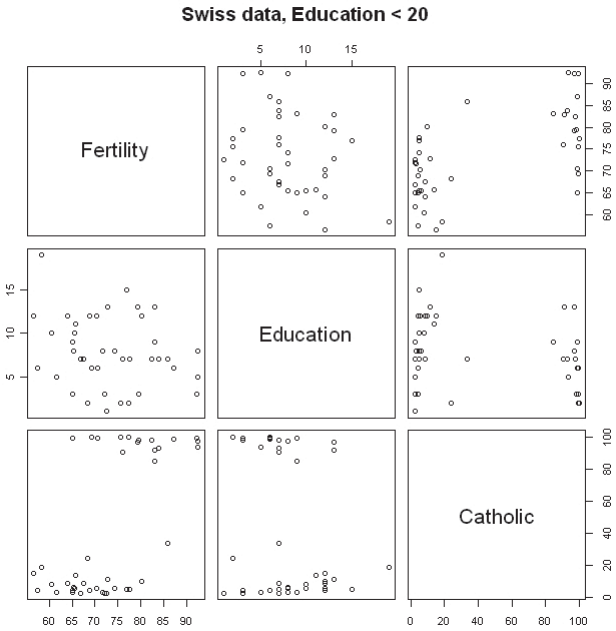
hist(sqrt(islands), breaks = 12, col=“lightblue”, border=“pink”)# **histograma**



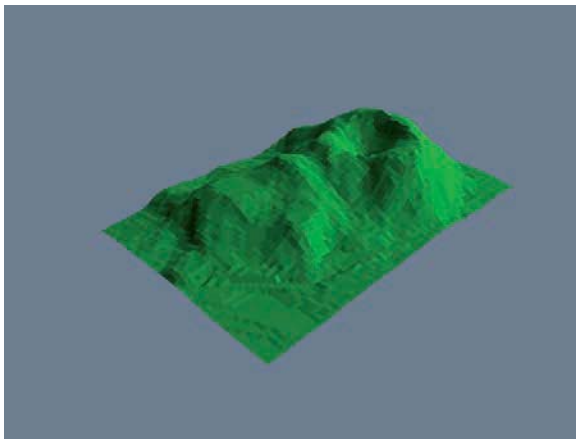
```
par(opar); opar <- par(pty = "s", bg = "lightcyan")
  plot(x = 0, y = 0,type = "n", xlim = rx, ylim = ry, xlab = "", ylab = "")
  u <- par("usr")
  rect(u[1], u[3], u[2], u[4], col = tcol[8], border = "red")
  contour(x, y, volcano, col = tcol[2], lty = "solid", add = TRUE,
      vfont = c("sans serif", "plain")))#gráfico de contorno
  title("A Topographic Map of Maunga Whau", font = 4)
  abline(h = 200*0:4, v = 200*0:4, col = "lightgray", lty = 2, lwd = 0.1
```




```
pairs(~ Fertility + Education + Catholic, data = swiss,
      subset = Education < 20, main = "Swiss data, Education < 20")#
```
pares de gráficos de dispersión por variables

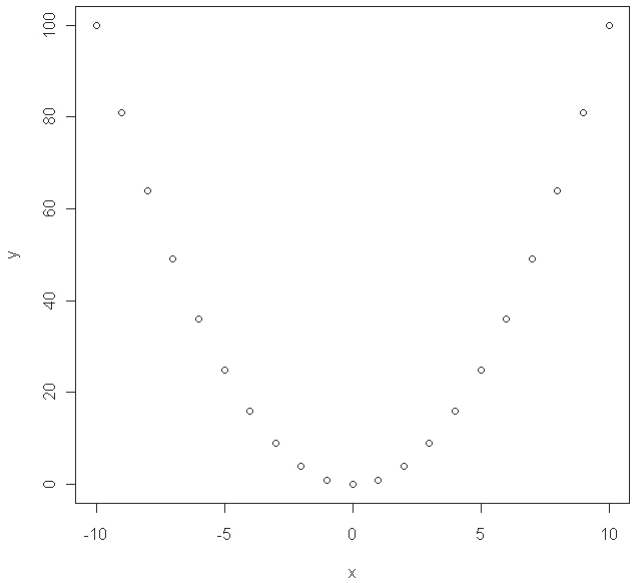



```
z <- 2 * volcano    # exagera el relieve
  x <- 10 * (1:nrow(z))  # 10 metros expansión (S a N)
  y <- 10 * (1:ncol(z))  # 10 metros expansión (E a O)
  ## No sacar las líneas de la cuadrícula: border = NA
  par(bg = "slategray")
  persp(x, y, z, theta = 135, phi = 30, col = "green3", scale = FALSE,
    ltheta = -120, shade = 0.75, border = NA, box = FALSE) #gráfico 3D
```



El comando más utilizado en gráficos es plot.

```
x<-c(-10:10)
y<-x^2
plot(x,y)
```



Se le puede añadir a plot varios argumentos:

axes= F	Suprimir los ejes
log='<x|y|xy>'	Ejes Logarítmicos
main='título'	Título
new=<logical>	Adiciona sobre el gráfico actual
sub='título de abajo'	Subtítulo
type='<l|p|b|n>'	Línea, puntos, ambos, ninguno
lty=n	Tipo de Línea
pch='.'	Carácter de dibujo
col=2	Para dibujar en color rojo
xlab='Nombre del eje x'	
ylab='Nombre del eje y'	
xlim=c(xminimo; xmaximo)	
ylim=c(yminimo; ymaximo)	

Gráficos de bajo nivel

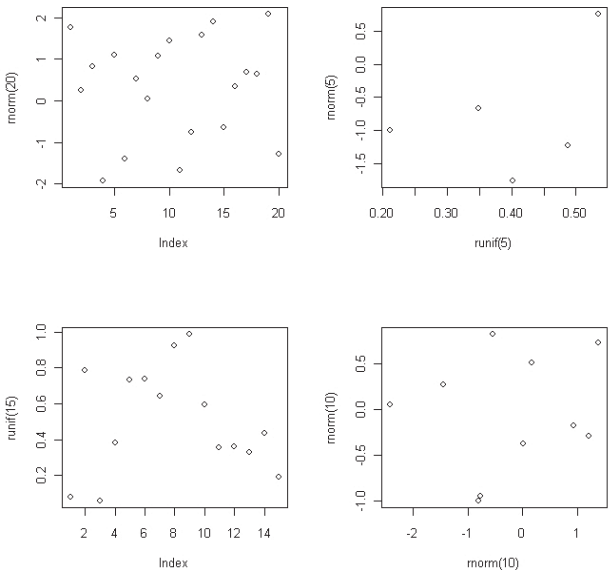
Se utilizan ciertos comandos para añadir información a los gráficos de alto nivel:

```
plot(x, y, type="p", lty=1:5, pch=, col=1:4)
points(x, y, type="p", lty=1:5, pch=, col=1:4) # añade puntos
lines(x, y, type="l", lty=1:5, pch=, col=1:4) # añade líneas
text(x,y, texto) # añade texto al gráfico en la posición x,y
abline(a, b) # añade una línea de pendiente a y de ordenada en el origen b
abline(h=y) # añade una línea horizontal
abline(v=x) # añade una línea vertical
qqplot(x, y, plot=TRUE)
qqnorm(x, datax=FALSE, plot=TRUE)
```

Múltiples gráficos por página

Empezamos con par(mfrow=c(filas,columnas)) antes del comando plot.

```
> par(mfrow = c(2, 2))
> plot(rnorm(20))
> plot(runif(5), rnorm(5))
> plot(runif(15))
> plot(rnorm(10), rnorm(10))
```



Guardar los gráficos

Se puede especificar donde guardar un gráfico.

```
> pdf(file = “graf1.pdf”, width = 10, height = 10)
> plot(rnorm(20))
> dev.off()
```

O bien, se puede copiar una figura a un fichero.

```
> plot(runif(50))
> dev.copy2eps()
```

10. ESTRUCTURA DE PROGRAMACIÓN Y FUNCIONES

10.1. Definición de función

Una función se define con una asignación de la forma

```
> nombre <- function(arg1,arg2,...){expresión}
```

La expresión es una fórmula o grupo de fórmulas que utilizan los argumentos para calcular su valor. El valor de dicha expresión es el valor que proporciona R en su salida y éste puede ser un simple número, un vector, una gráfica, una lista o un mensaje.

Ejemplo: Suma de una progresión aritmética

```
> suma <- function(a1,d,n){an <- a1+(n-1)*d;((a1+an)*n)/2}
```

Cuya ejecución se realiza:

```
> suma <- function(a1,d,n){an <- a1+(n-1)*d;((a1+an)*n)/2}
> suma(3,4,20)
[1] 820
```

10.2. Control de ejecución

Principales instrucciones:

if(cond) expr # Si se cumple la condición se hace la exp

if(cond) cons.expr else alt.expr # Si se cumple la condición se hace cons.expr en caso contrario alt.expr

for(var in seq) expr # Hacer según la variable índice var hasta seq la expresión expr

while(cond) expr # Hacer mientras se cumpla la condición la expresión expr

repeat expr # Repetir expr

break # Salir del bucle

next # Continuar

11. LA EXPLORACIÓN DE DATOS: ESTADÍSTICOS Y FUNCIONES DE DISTRIBUCIÓN DE PROBABILIDAD

11.1. La exploración de los datos

Para estudiar la distribución de los datos se puede realizar con las funciones ***summary*** o ***fivenum***; y también se puede realizar un diagrama de tallo y hojas con la función ***stem***..

```
> data(faithful)
> attach(faithful)
> summary(eruptions)
```

```
> summary(eruptions)
 Min. 1st Qu. Median  Mean 3rd Qu.  Max.
1.600  2.163  4.000  3.488  4.454  5.100> stem(eruptions)
```

```
> stem(eruptions)
```

```
> stem(eruptions)

The decimal point is 1 digit(s) to the left of the |

16 | 070355555588
18 | 00002223333333557777777788822335777888
20 | 00002223378800035778
22 | 0002335578023578
24 | 00228
26 | 23
```

```
28 | 080
30 | 7
32 | 2337
34 | 250077
36 | 0000823577
38 | 2333335582225577
40 | 00000033577888880022335555777 78
42 | 03335555778800233333555577778
44 | 022223355577800000000233333577 78888
46 | 0000233357700000023578
48 | 00000022335800333
50 | 0370
```

También se puede realizar un contraste para comprobar la normalidad.

```
> long <- eruptions[eruptions > 3]
> shapiro.test(long)
```

```
> shapiro.test(long)

    Shapiro-Wilk normality test

data: long
W = 0.9793, p-value = 0.01052
```

y el contraste de Kolmogorov-Smirnov

```
> ks.test(long, "pnorm", mean=mean(long), sd=sqrt(var(long)))
```

```
> ks.test(long, "pnorm", mean=mean(long), sd=sqrt(var(long)))

            One-sample Kolmogorov-Smirnov test

data: long
D = 0.0661, p-value = 0.4284
alternative hypothesis: two-sided
```

11.2. Distribuciones de probabilidad

Existen funciones de librería para distintas distribuciones de probabilidad:

Gausse (normal)	rnorm(n, mean=0, sd=1)
exponencial	rexp(n, rate=1)
gamma	rgamma(n, shape, scale=1)
Poisson	rpois(n, lambda)
Weibull	rweibull(n, shape, scale=1)
Cauchy	rcauchy(n, location=0, scale=1)
beta	rbeta(n, shape1, shape2)
'Student' (t)	rt(n, df)
Fisher–Snedecor (F)	rf(n, df1, df2)
Pearson (c2)	rchisq(n, df)
binomial	rbinom(n, size, prob)
geométrica	rgeom(n, prob)
hypergeométrica	rhyper(nn, m, n, k)
logística	rlogis(n, location=0, scale=1)
lognormal	rlnorm(n, meanlog=0, sdlog=1)
binomial negativa	rnbinom(n, size, prob)
uniforme	runif(n, min=0, max=1)
Estadístico de Wilcoxon's	rwilcox(nn, m, n), rsignrank(nn, n)

11.3. Funciones para estadísticas simples

cor	Correlación (admite uno o dos argumentos)
cumsum	Suma cumulativa de un vector
mean	Media aritmética
median	El percentil 0.5: la mediana
min	El mínimo de una serie de números
max	El máximo de una serie de números
prod	El producto de los elementos de un vector
quantile	Los percentiles de una distribución
range	Mínimo y máximo de un vector
sample	Muestreo aleatorio (y permutaciones)
sum	Suma aritmética
var	Varianza y covarianza
summary	Resumen de estadísticas de una serie de datos

12. DIVERSOS CONTRASTES ESTADÍSTICOS

t.test(x,y=NULL, alternative=”two.sided”, mu=0,paired=F,var.equal=T, conf.level=.95)	Realiza el contraste de hipótesis de la media de x o de la diferencia de medias (x-y)de una población normal.
binom.test(x,n,p=0.5, alternative=”two.sided”)	Test exacto para una proporción, x=número de éxitos, n=número de pruebas.
prop.test(x,n,p, alternative=”two.sided”, conf.level=0.95,correct=T)	Test de la diferencia de proporciones,x=número de elementos de las muestras que cumplen una característica, n=vector del número de elementos en las muestras, p=vector de probabilidad de éxito.
var.test(x,y, alternative=”two.sided”)	Contraste del cociente de varianzas x e y
ks.gof(x,y=NULL, distribution=”normal”, alternative=”two.sided”)	Test de Kolmogorov-Smirnov para ajuste de datos a distribuciones dadas. Habrá que cargar librería gldex
shapiro.test	Test de Shapiro para comprobar ajuste de datos a una distribución normal. Habrá que cargar librería cwhmisc
runs.test as.factor(x>median(x)))	Aplica el test de rachas, pero previamente habrá que cargar library(“tseries”). Con as.factor pasa la variable a dicotómica en función del valor mediana de x
chisq.test(x,y=NULL, correct=T)	Test de la chi-cuadrado para datos de conteo, x=es una matriz o tabla de contingencia
fisher.test(x,y=NULL,..)	Test exacto de Fisher para datos de conteo, x=es una matriz o tabla de contingencia

wilcox.test(x,y, alternative="two.sided", mu=0,paired=F,exact=F, correct=T)	Para una única muestra x, realiza el test de Wilcoxon de rangos con signos. Para dos muestras x e y y además paired=F, realiza el test de Mann-Whitney para dos muestras independientes. Si paired=T, realiza el test de Wilcoxon de rangos con signos para muestras relacionadas
Kruskal.test(y,groups)	Test de Kruskal-Wallis, y=vector numérico y groups=diferentes grupos para contrastar la hipótesis de que todos ellos proceden de la misma población
friedman.test(y,groups ,blocks)	Test de Friedman, y=vector numérico y groups=diferentes tratamientos y blocks=diferentes muestras extraídas
cor.test(x, y, alternative = c("two.sided", "less", "greater"), method = c("pearson", "kendall", "spearman"), exact = NULL, conf.level = 0.95, ...)	Test de correlación según las características de los datos
aov(formula,data)	Calcula el análisis de la varianza de los datos según la fórmula indicada
lm(formula,data)	Ajusta a un modelo lineal a los datos especificados según la fórmula. Por defecto incorpora ordenada en el origen. Si no interesa añadir -1 en la fórmula. La fórmula tampoco incluye los coeficientes del modelo. La fórmula de la regresión simple será:y~x

BIBLIOGRAFÍA

Anastasi, A. (1980). *Tests psicológicos*. Madrid: Aguilar.

Angulo Rasco, J. F. (1990). "Una propuesta de clasificación de las técnicas de recogida de información, con especial referencia a las técnicas de investigación interpretativas". En J. B. Martínez Rodríguez (ed.), *Hacia un enfoque interpretativo de la enseñanza* (pp. 85-93). Granada: Servicio de Publicaciones de la Universidad de Granada.

Bisquerra, R. (1989). *Métodos de Investigación educativa. Guía práctica*. Barcelona: Ceac.

Bugeda, J. (1975). *Sociología matemática.* Madrid: Instituto de Estudios Políticos.

Bunge, M. (1976). *La investigación científica*. Barcelona: Ariel.

Calvo, F. (1990). *Estadística aplicada: con el planteamiento y resolución de 450 problemas.* Bilbao: Deusto.

Carmines, E. G. y Zeller, R. A. (1979). Reliability and validity assessment. Londres: Sage.

Coombs, C. H.; Dawes, R. M. y Tversky, A. (1981). *Introducción a la psicología matemática*. Madrid: Alianza Universidad.

Cronbach, L. J. (1972). *Fundamentos de la exploración psicológica*. Madrid: Biblioteca Nueva.

De Ketele, J. M. y Roegiers, X. (1995). *Metodología para la recogida de información*. Madrid: La Muralla.

Del Rincón, D. y otros (1995). *Técnicas de investigación en Ciencias Sociales*. Madrid: Dykinson.

Duverger, M. (1974). *Método de las ciencias sociales*. Barcelona: Ariel.

Evertson, C. M. y Green, J.L. (1986). "Observation as inquiry and method". En M. C. Wittrock (ed.), *Handbook- of research on teaching* (pp. 162-213). New York: MacMillan Pub. Co.

Fontana, A. y Frey, J. H. (1994). "Interviewing: The Art of Science". En N, K. Denzin y Y. S. Lincoln (eds.), *Handbook of Qualitative Research* (pp. 361-376). Londres: Sage.

García Jiménez, E. (1994). "Investigación etnográfica". En V. García Hoz (dir.), *Problemas y métodos de investigación en Educación Personalizada* (pp. 343-375). Madrid: Rialp.

Gaviria Soto, J.L. (2000). "Fundamentos de la medida en Educación y Psicología. Del representacionalismo a la medida como hipótesis". *Revista de Ciencias de la Educación:* Organo del Instituto Calasanz de Ciencias de la Educación (n.º 184, pp. 207-224). Madrid.

Kerlinger, F. N. (1985). *Investigación del comportamiento. Técnicas y metodología*. México: Interamericana.

Martínez Arias, R. (1995). *Psicometría: teoría de los tests psicológicos y educativos*. Madrid: Síntesis.

Muñiz, J. (2000). *Teoría clásica de los tests.* Madrid: Pirámide.

Muñiz, J. (1997). *Introducción a la Teoría de respuesta a los ítems*. Madrid: Pirámide.

Morales, P. (1988). *Medición de actitudes en psicología y educación. Construcción de escalas y problemas metodológicos*. San Sebastián: Tartalo.

Osgood, G.; Suci, G. y Tannembaum, P. (1976). *La medida del significado.* Madrid: Gredos.

Pérez Juste, R. (1986). *Pedagogía experimental. La medida en educación*. Madrid: UNED.

Siegel, S. (1970). *Estadística no paramétrica. Aplicada a las ciencias de la conducta.* México: Trillas.

Summers, G. F. (1976*). Medición de actitudes*. México: Trillas.

Walker, R. (1989). *Métodos de investigación para el profesorado.* Madrid: Morata.

Capítulo 2
CARACTERÍSTICAS TÉCNICAS DE LOS INSTRUMENTOS DE MEDIDA

1. Características técnicas de los instrumentos de medida
2. La teoría clásica de los test (TCT)
   2.1. Dificultad
   2.2. Discriminación
   2.3. Análisis de los distractores
   2.4. Fiabilidad
   2.5. Validez
   2.6. Dimensionalidad
3. Teoría de respuestas al ítem (TRI)
   3.1. Supuestos
   3.2. Modelos
   3.3. Curva característica del test
   3.4. Funciones de información
4. Aplicaciones de la TRI
5. Un ejemplo de cálculo de las características técnicas de un test

Objetivos

- Comprender la utilidad de la medición en educación
- Saber calcular la dificultad, discriminación, análisis de los distractores, fiabilidad y validez de un test
- Saber calcular la unidimensionalidad de un test: sólo como ampliación
- Diferenciar la TCT de la TRI
- Calcular, mediante R, los parámetros de un ítem desde la TRI
- Interpretar la información de los parámetros y características de un ítem desde la TRI

1. CARACTERÍSTICAS TÉCNICAS DE LOS INSTRUMENTOS DE MEDIDA

El proceso de medición opera sobre un ***objeto***, según unas ***reglas*** —que definen el isomorfismo entre el objeto y la ***unidad*** de medida—, y con una unidad de medida. Este proceso nos lleva a construir unos ***instrumentos***, que se pasan al objeto en el ***acto de medición***, teniendo como resultado unos ***datos*** con los que se realizará un ***tratamiento***, generalmente estadístico.

En este capítulo se hablará de las características técnicas de los instrumentos de medida, sobre todo en lo referente a los **ítems** —componentes de cuestionarios, tests y pruebas objetivas- según un planteamiento clásico -lo que se ha venido en llamar la teoría clásica de los tests—, cuyo resultado depende de la muestra elegida, fundamentalmente en el aspecto de variabilidad de la misma y son expresados en relación al "grupo normativo". Sólo en algunos momentos —coeficiente kappa— se hará mención a los test de referencia criterial, aunque dicho coeficiente no es de uso exclusivo en dichos test. La segunda parte del capítulo se dedicará a otro modelo de medida basado en la llamada teoría de respuesta al ítem o del rasgo latente, cuyos resultados son invariantes a la muestra elegida aunque exige un importante tamaño de la misma.

2. TEORÍA CLÁSICA DE LOS TEST (TCT)

Son muchas las características observables en los ítems, no obstante se verán aquellos aspectos relacionados con las propiedades de los test. Para evaluar la calidad de los ítems los indicadores más utilizados son:

- Dificultad
- Discriminación

- Análisis de los distractores
- Fiabilidad
- Validez
- Dimensionalidad

2.1. Dificultad

El grado de dificultad de un ítem está en función del número de personas que contesten de forma correcta al mismo. El índice de dificultad es un indicador de la misma.

$$ID = \frac{A}{N}$$

donde:

A = número de personas que aciertan el ítem

N = número total de personas que responden al ítem.

En ítems de elección múltiple, para corregir los efectos del azar, se suele utilizar el llamado índice de dificultad corregido:

$$ID' = p - \frac{q}{k-1} = \frac{A - \frac{E}{k-1}}{N}$$

donde:

p = proporción de aciertos (índice de dificultad sin corregir)

q = proporción de fallos

k = número de alternativas del ítem

o también:

A = aciertos

E = errores

N = total

El ID no es una propiedad intrínseca del ítem, su valor depende de la muestra de sujetos a la que se aplique.

Naturalmente el índice de dificultad tiene sentido en test o pruebas objetivas de medida de aptitudes o rendimiento donde es relevante decir si una respuesta es correcta o no.

Los valores del índice de dificultad de un ítem se suelen clasificar en cinco categorías:

Muy fáciles	ID por encima de 0,75
Fáciles	ID comprendido entre 0,55 y 0,75
Normales	ID comprendido entre 0,45 y 0,54
Difíciles	ID comprendido entre 0,25 y 0,44
Muy difíciles	ID por debajo de 0,25

En la construcción de test es aconsejable poner mayor cantidad de ítems con una dificultad media (aquellos con histograma simétrico) y una pequeña cantidad de ítems con dificultad alta y otra con dificultad baja. Entiéndase que el índice de dificultad (en realidad se debería llamar índice de facilidad) es una función creciente que va desde valores bajos, cuando la dificultad es mayor, hasta valores altos (por ejemplo 0,9 cuando responden acertadamente el 90% de las personas).

Ejemplo 2.1

a) Se supone un ítem con dos alternativas de una prueba que contestan acertadamente 25 estudiantes en una clase de 40 alumnos y otro ítem que sólo aciertan 15 personas. Los índices de dificultad serán:

$$ID = \frac{25}{40} = 0,625 \text{ que según los valores anteriores se considerará fácil}$$

$$ID = \frac{15}{40} = 0,375 \text{ que se considerará difícil}$$

b) Ahora se tienen los resultados de un ítem de tres alternativas donde aciertan 25 estudiantes en una clase de 40 alumnos y otro ítem con 15 resultados acertados. Los índices de dificultad serán:

$$ID' = \frac{25 - \frac{15}{(3-1)}}{40} = 0,4375 \text{ que será considerado difícil}$$

$$ID' = \frac{25 - \frac{25}{(3-1)}}{40} = 0,0625 \text{ que será considerado muy difícil}$$

2.2. Discriminación

Este índice determina la selectividad de la pregunta para distinguir o diferenciar el grupo que tiene mayor cantidad de aciertos en la misma, frente al de menos aciertos. Es decir, un ítem discrimina cuando existe correlación positiva entre la puntuación obtenida en el ítem con la puntuación obtenida en el test, desde este punto de vista el índice de discriminación da una idea de la homogeneidad del ítem en relación al test. Este índice va de –1 a +1, considerándose un buen índice de discriminación el comprendido entre 0,25 y 0,35 o mayor.

$$D = p_+ - p_-$$

donde:

Dividida la muestra en dos grupos extremos: a) una parte donde está el 27% de personas con mayor puntuación en el test (extremo superior) y b) otra parte con el 27% de menor puntuación (extremo inferior).

p_+ representa la proporción de individuos del extremo superior que aciertan el ítem

p_- representa la proporción de individuos del extremo inferior que aciertan el ítem

Otra manera de valorar la discriminación de un ítem es mediante el **cálculo de la correlación obtenida entre las puntuaciones del grupo en el ítem y en el test**, cuando este ítem no es valorado.

Esta correlación se denomina **índice de discriminación** o de **homogeneidad** al ser un indicador del grado de relación del ítem con el resto. Si la

correlación es elevada las personas que puntúan alto en el test acertarán también en el ítem. Si no se quita el ítem, a la hora de realizar los cálculos, la fórmula a utilizar es la siguiente:

$$IH = \frac{r_{ix}S_x - S_i}{\sqrt{S_x^2 + S_i^2 - 2r_{ix}S_xS_i}}$$ donde: IH de discriminación o de homogeneidad

r_{ix} es la correlación ítem-test

S_x es la desviación típica del test

S_i es la desviación típica del ítem

La correlación ítem-test se calculará con diferente coeficiente de correlación en función del tipo de distribución conjunta de las variables que se correlacionan: a) si son dicotómicas el coeficiente ϕ (correlación de Pearson de dos variables dicotómicas), b) si están dicotomizadas y se suponen provienen de distribuciones normales, la correlación tetracórica; c) una variable continua y otra dicotomizada, correlación biserial; d) una variable continua y otra dicotómica, correlación biserial puntual (aplicación de la correlación de Pearson entre una variable dicotómica y otra cuantitativa); y f) dos variables continuas, correlación de Pearson.

Algunos autores fijan los siguientes valores a la hora de calificar el grado de homogeneidad:

El ítem discrimina muy bien	IH igual o mayor de 0,4
El ítem discrimina bien	IH comprendido entre 0,3 y 0,39
El ítem discrimina poco	IH comprendido entre 0,20 y 0,29
Ítem límite. Se debe mejorar	IH comprendido entre 0,10 y 0,19
El ítem carece de utilidad para discriminar	IH menor de 0,10

Variabilidad y discriminación

La variabilidad en el test está relacionada con el índice de discriminación:

$$S_x = \sum_{j=1}^{n} S_j r_{jx}$$

donde:

S_x = desviación típica del test

S_j = desviación típica del ítem

r_{jx} = índice de discriminación del ítem j

Ejemplo 2.2

a) Se supone que un ítem de un test se pasa a 300 estudiantes de los 81(27% de 300) que obtienen mejores puntuaciones en el test, 40 aciertan el ítem y de los 81 que obtienen peores notas en el test, 12 aciertan el ítem, en consecuencia:

$$IH = \frac{40}{81} - \frac{12}{81} = 0,3456 \text{ en consecuencia el ítem discrimina bien}$$

b) Se realiza el estudio de ítems de una prueba de 15 ítems medidos en una escala continua. Los resultados de alguno de los ítems son los siguientes:

	Correlación elemento-total corregida
MAT4	,383
MAT5	,568
MAT6	,386
MAT7	,570
MAT8	,647
MAT9	,399
MAT10	,533
MAT11	,092
MAT12	,420
MAT13	,519
MAT14	,529
MAT15	,382

Según se puede observar el ítem MAT11 tiene un índice de discriminación bajo (0,092) y por tanto se debería quitar de la prueba.

2.3. Análisis de distractores

Distractores son las diferentes alternativas falsas o respuestas incorrectas que tiene un ítem. Todas las alternativas de un ítem deberían tener la misma probabilidad de ser elegidas, es decir, se debería dar la equiprobabilidad entre las diferentes alternativas de respuesta. Para comprobar tal circunstancia se puede plantear una tabla donde figuren las frecuencias de elección de cada alternativa y luego utilizar la prueba de ji-cuadrado y así comprobar la independencia de las elecciones.

Ejemplo 2.3

Se supone que 300 individuos contestan a un ítem con cuatro alternativas y que la D es la respuesta correcta. Los valores obtenidos para los 198 individuos que fallan en las distintas alternativas erróneas son los siguientes:

A	B	C
85	40	73

$$\chi^2 = \frac{(85-66)^2 + (40-66)^2 + (73-66)^2}{66} = \frac{1086}{66} = 16,45$$

El estadístico anterior se distribuye según una ji-cuadrado con k-1 grados de libertad. Donde k es el número de categorías incorrectas (en el ejemplo k=3). Por tanto a un nivel de confianza del 95%, ji-cuadrado teórico=5,991. Por tanto se rechaza la hipótesis de independencia y las respuestas son distintas en función de la alternativa.

2.4. Índice de fiabilidad

Supongamos el modelo lineal clásico de medida (Muñiz, 2000 p. 28):

Modelo: X=V+e

Supuestos:

1. E(X)=V, la media de las puntuaciones empíricas es igual a la puntuación verdadera
2. ρ(v,e)=0, la correlación entre la puntuación verdadera y el error de medida es 0
3. $\rho(e_j,e_k)=0$, la correlación entre los errores de medida es 0

donde X es la puntuación empírica, V la puntuación verdadera y e el error de medida cometido en la medición.

Un instrumento de medida es fiable si está libre de error de medida. Aunque un instrumento de medida no es bueno por el hecho de ser fiable, es decir, de ofrecer unos números merecedores de confianza, o sea, libres de errores de medida, sin tal propiedad la medición no es aceptable en la investigación; esto es, podemos afirmar que la fiabilidad es una ***condición necesaria***, aunque no suficiente, de un instrumento de medida de calidad.

La fiabilidad de un instrumento de medida se cuantifica mediante el coeficiente de fiabilidad (fiabilidad relativa) y con el error típico de medida (fiabilidad absoluta).

El coeficiente de fiabilidad $\rho_{XX'}$ se define como la correlación entre las puntuaciones obtenidas por los sujetos en dos formas paralelas de un test, X y X' (Muñiz 2000, p.35).

La fiabilidad es un indicador de la probabilidad de fallo que tiene nuestro instrumento en la reproducción de la medida o en la utilización de un instrumento paralelo, complementado por el grado de discrepancia que tiene respecto al verdadero valor.

De los supuestos del modelo lineal se deduce:

$$\rho_{XX'} = \frac{\sigma_V^2}{\sigma_X^2} \tag{1}$$

y como var(X)=var(V)+var(e):

$$\rho_{XX'} = 1 - \frac{\sigma_e^2}{\sigma_X^2} \tag{2}$$

Error típico de medida

Si nos fijamos en el segundo aspecto de la fiabilidad la ***precisión***, tendremos que hablar de la ***fiabilidad absoluta*** definida en términos del ***error típico de medida***.

Si repetimos indefinidamente un test a un individuo, la media de las puntuaciones nos indicará el verdadero valor de la puntuación, de igual forma la desviación típica de las puntuaciones respecto de la media verdadera es el error típico.

Como es impensable repetir el test a una persona indefinidamente tendremos que encontrar una estimación. Operando en la fórmula (2) tenemos la expresión del error típico de medida:

$$\sigma_e = \sigma_x \sqrt{1 - \rho_{xx'}}$$

Las aplicaciones del error típico de medida son:

— Estimación de los límites entre los que se encuentra la verdadera medida.

— Comparar las puntuaciones de varios sujetos en el mismo instrumento o del mismo sujeto en diferentes instrumentos.

Como los valores de la varianza del error y la varianza de la puntuación verdadera son desconocidos será necesario realizar estimaciones.

Estimaciones empíricas del coeficiente de fiabilidad

* *FIABILIDAD COMO ESTABILIDAD*

Nos fijamos en la constancia de las puntuaciones en un sentido temporal. Dos aplicaciones del mismo instrumento separadas por un tiempo variable.

Este tipo de fiabilidad se ve influido por las condiciones cambiantes de carácter de los individuos sometidos a la prueba.

Se mide con el llamado ***coeficiente de estabilidad***, cuyo procedimiento de obtención es el siguiente: se realizan dos aplicaciones espaciadas (test,

retest) del instrumento de medida y luego se calcula el coeficiente de correlación entre las puntuaciones obtenidas en las mismas.

La fórmula utilizada para calcular el coeficiente de fiabilidad entre las dos mitades es el procedimiento de Spearman-Brown:

$$R_{xx} = \frac{2.r_{xx}}{1+r_{xx}}$$

donde r_{xx} es el coeficiente de correlación de Pearson entre las mitades. Así por ejemplo si r_{xx} = 0,84 el coeficiente de fiabilidad será:

$$R_{xx} = \frac{2.0,84}{1+0,84} = 0,913$$

Existen dos problemas que pueden influir claramente en los resultados: la memoria de los individuos ante la repetición de la prueba y la motivación de los mismos (desgana o desinterés ante la reproducción del evento). El tiempo fijado por diversos autores entre las dos aplicaciones del instrumento está entre 20 a 25 días.

* *FIABILIDAD COMO EQUIVALENCIA*

Un instrumento es equivalente a cualquier otro que mida el mismo aspecto que pretendemos medir, es decir, suponemos que "el instrumento es una muestra suficiente y representativa de la población de conductas que integran el rasgo o característica a medir" (Pérez Juste, 1986).

Los errores de medida que se puedan producir son debidos a falta de equivalencia entre los elementos muestreados o a experiencias de los sujetos anteriores a la realización de la prueba.

Este aspecto de la fiabilidad se mide con el llamado ***coeficiente de equivalencia***. El procedimiento para su obtención es el siguiente: se toman dos muestras de una misma población, es decir dos conjuntos de ítem (dos formas paralelas) que, según Pérez Juste (1986), deben reunir las siguientes características:

1. Deben tener el mismo número de elementos.
2. La redacción y la estructura de cada elemento paralelo debe ser la misma.

3. El contenido y el objetivo apreciado debe ser el mismo.
4. Los índices de dificultad deben ser iguales.
5. No debe existir diferencia significativa entre medias, varianzas y covarianzas.
6. Las consignas que se den en la presentación deben ser las mismas, para que ninguno resulte más atrayente que el contrario.
7. Presentación idéntica.

Con estas sub-muestras o instrumentos equivalentes se hacen, como en el caso anterior, dos aplicaciones, una con cada una de las formas, viniendo determinado el coeficiente de equivalencia por el coeficiente de correlación obtenido con las puntuaciones de ambas aplicaciones.

En el planteamiento de las pruebas de referencia criterial, uno de los coeficientes de fiabilidad más utilizado en formas paralelas es el coeficiente k (kappa):

$$K = \frac{F_c - F_a}{N - F_a}$$

donde F_c es la frecuencia para los elementos que coinciden y F_a es la frecuencia de azar para los elementos que coinciden y N, número de casos. El cálculo de este coeficiente es similar al coeficiente de correlación de Pearson para datos dicotómicos. Indudablemente este coeficiente tendrá como valor máximo 1 y como mínimo 0.

* *Fiabilidad como consistencia interna*

El índice de fiabilidad (IF) de un ítem es función del índice de discriminación y de la desviación típica del ítem

$$IF = S_i D_i$$

donde:

S_i es la desviación típica en el ítem i

D_i índice de discriminación del ítem i

Cabe pensar que todos los elementos del instrumento de medida miden una porción de la característica o rasgo, de forma que debe darse una

coherencia interna, es decir una sub-muestra de ítems mide lo mismo que otra sub-muestra, y representan ambas lo que mide el instrumento en su totalidad.

El grado de correlación entre ambas sub-muestras determina el ***coeficiente de consistencia interna***. Para formar estas sub-muestras se agrupan en una los ítem pares y en otra los impares, o bien la primera mitad y la segunda, aunque esta segunda opción tiene el inconveniente de la distribución uniforme en dificultad del test.

Al utilizar sub-muestra, el coeficiente de correlación será necesario corregirlo para obtener la fiabilidad de todo el test.

Para calcular la fiabilidad del test completo, como consistencia interna de la escala, se utiliza el estadístico α de Cronbach, estimador por defecto del coeficiente de fiabilidad, es decir $\rho_{XX'} \geq \alpha$.

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\sum_{j=1}^{n} S_j^2}{S_x^2}\right)$$

Resulta del cociente entre las suma de las varianzas de los ítems, dividida por la varianza total, ponderada por el número de ítems de la escala.

Como vemos, la fiabilidad de un test está relacionada directamente con el número de ítems del mismo: a mayor cantidad de ítems más fiabilidad; por eso, de alguna forma, esta fórmula nos sirve de elemento predictor de la fiabilidad cuando se aumenta la cantidad de ítems.

Según apunta Muñiz (2000, p. 55) un elevado valor de alfa no es un indicador de la unidimensionalidad de la escala de medida, sino que debe completarse con otras técnicas.

Si el cuestionario o test es multidimensional o multifacético no es correcto medir la consistencia interna de todos, sino de las diferentes facetas o subescalas por separado.

Ejemplo 2.4

Se supone el conjunto de resultados del ejemplo 2.2 apartado b) los estadísticos de los ítems son:

```
     mean        sd        n  NA
MAT1  1.0000000 0.0000000 59  2
MAT2  0.5254237 0.5036396 59  2
MAT3  0.8474576 0.3626321 59  2
MAT4  0.6610169 0.4774274 59  2
MAT5  0.7627119 0.4290721 59  2
MAT6  0.9152542 0.2808936 59  2
MAT7  0.5254237 0.5036396 59  2
MAT8  0.5593220 0.5007300 59  2
MAT9  0.8644068 0.3452948 59  2
MAT10 0.7457627 0.4391693 59  2
MAT11 0.6101695 0.4918981 59  2
MAT12 0.8644068 0.3452948 59  2
MAT13 0.6724138 0.4734321 58  3
MAT14 0.5517241 0.5016609 58  3
MAT15 0.7068966 0.4591610 58  3
```

Por tanto si se quiere calcular la fiabilidad del ítem 7, por ejemplo, el resultado será:

$$IF = 0{,}5036*0{,}570 = 0{,}287$$

Si se quiere calcular el índice de fiabilidad como consistencia interna del conjunto de ítems, es decir, de la escala, el valor del α de Cronbach, dado el volumen de cálculo, se debe utilizar algún programa estadístico como R. En el apartado 5 se abordará un ejemplo completo con R.

2.5. Validez

Como hemos comentado, la fiabilidad de un instrumento, es condición necesaria pero no suficiente para una valoración positiva de éste; necesita el concurso de la validez para completar dicha valoración.

La idea de validez es un concepto relativo. Para Pérez Juste (1986) un instrumento es válido si mide lo que dice medir. Por su parte Félix Calvo (1990) nos argumenta que la validez puede ser exigida respecto a:

— La ***construcción*** del test; para ello se han de estudiar si las cualidades o factores que queremos medir están valorados en ese test.

— El ***contenido***: las cuestiones que debe abarcar el contenido de lo que se quiere medir.

— El ***criterio***, que se logra con otras puntuaciones externas al test, de modo que con ellas, y a través de la correlación con las puntuaciones del test, pueda detectarse si éstas son válidas.

Según la American Psychological Associatión, hay cuatro metas de los instrumentos de medición que representan cuatro tipos de validez: ***contenido, predictiva, concurrente y de constructo***; otros autores incluyen la ***aparente***.

VALIDEZ DE CONTENIDO

Es imprescindible sobre todo en pruebas de rendimiento. De ella nos dice Pérez Juste (1986, p. 227): "La esencia de la validez de contenido se encuentra en la suficiencia y representatividad de la muestra de conductas incluidas en el instrumento respecto de la población de la que forma parte".

La suficiencia es un problema de tamaño, ya que necesitamos buscar la longitud adecuada de nuestro instrumento, para que por una parte incorpore todas las facetas que queremos medir y, por otra, no resulte, en modo alguno, aburrido y penoso para el evaluado. La representatividad es obvio que resulta imprescindible para los instrumentos de medida.

VALIDEZ PREDICTIVA

La predicción, de forma genérica, es uno de los objetivos de todo trabajo científico. Siempre necesitamos saber la capacidad de inferencia que posee nuestro instrumento de medida.

La forma de obtener un indicador de esta predicción, es determinar la correlación entre las puntuaciones obtenidas por los sujetos en la prueba con alguna otra medida.

Se necesitan varias condiciones para realizar con éxito esta comparación:

- Debe mediar algún tiempo entre las aplicaciones.
- Debe existir independencia entre ambas puntuaciones, es decir, los instrumentos no tienen que estar relacionados.

- Se necesitan criterios intermedios, relativos, para realizar las predicciones.

Cuando se desea realizar pronósticos sobre una variable externa (criterio) con el instrumento de medida construido, se necesita valorar el grado de relación de cada ítem con dicho criterio, esta relación se valorará con el llamado **índice de validez del ítem** que no es otra cosa que la correlación entre el ítem y el criterio.

$$IV = r_{jy}$$

Debemos tener presente que la validez predictiva no implica validez de contenido, ya que un instrumento puede, por ejemplo, predecir el éxito en una determinada profesión sin tener contenidos apropiados.

Validez concurrente

En realidad es una variante de la predictiva. Se diferencia de la misma en que las medidas se hacen simultáneamente con el fin de tomar decisiones en un plazo breve.

Igual que en el caso anterior, la validez concurrente no implica validez de contenido, aunque sí es conveniente que se den las dos.

Validez de elaboración o de constructo

Es llamada por algunos de hipótesis de trabajo. Según Pérez Juste (1986, p. 240) la validez de constructo permite ir a la raíz del problema: "eso que medimos ¿qué es?".

La cuantificación de esta validez requiere una acumulación progresiva de información que irá dando más consistencia a la construcción inicial, o por el contrario ira desechando el planteamiento presentado. El método utilizado para conseguir estos fines es el análisis factorial de las tareas incluidas en la prueba.

Validez y longitud de una prueba

Como se ha comentado anteriormente existe una relación directa entre fiabilidad y el número de ítems de la prueba. La validez también se relaciona con la longitud y con la fiabilidad de la prueba:

$$R_{xy} = \frac{r_{xy}}{\sqrt{\frac{1-r_{xx}}{n} + r_{xx}}}$$ con:

R_{xy} la nueva validez

r_{xy} la validez primitiva

r_{xx} la fiabilidad inicial

n cociente entre número de ítems iniciales y el número de ítems finales

2.6. Dimensionalidad

Los test tienen como objetivo medir una variable de interés. Esta variable se puede concretar en una o más dimensiones, por tanto uno de los aspectos más importante a medir en el test es la dimensionalidad del mismo. Para realizar dicho estudio la metodología más apropiada es el análisis factorial de los ítems, con objeto de observar la agrupación de los mismos en uno o más factores.

El análisis factorial es una técnica estadística de reducción de dimensión que trabaja con la matriz de correlaciones o covarianzas como entrada. Si los ítems son dicotómicos la matriz de correlaciones se construirá con las correlaciones tetracóricas y si los ítems son categóricos, por ejemplo una escala de likert de 1 a 5 categorías, la matriz de correlaciones más apropiada será las correlaciones policóricas.

Existen distintas técnicas de factorización para el agrupamiento de ítems, sin embargo las más utilizadas por su bondad son el método de componentes principales y el método de ejes factoriales. Este último es utilizado para maximizar la información de cada factor.

También hay diversos criterios para considerar la unimensionalidad cuando se obtienen diversos factores en el análisis factorial, dado por supuesto la imposibilidad de explicar el 100% de la varianza común. Quizás el más utilizado es el criterio de Carmines y Zeller (1979) que consideran la existencia de unidimensionalidad cuando el primer factor explica más del 40% de la varianza común. También según Lord (1980) se considera un test

unidimensional si el primer factor es, como mínimo, cinco veces mayor que el segundo, o de otra forma, si $\lambda_1/\lambda_2 > 5$ siendo λ_1 y λ_2 los autovalores de los dos primeros factores.

El cálculo de la dimensionalidad de un cuestionario o prueba se sale de los objetivos del texto, no obstante los lectores interesados pueden consultar libros sobre análisis multivariante como el nuestro.[1]

3. TEORÍA DE RESPUESTA AL ÍTEM (TRI)

La Teoría de Respuesta al ítem (TRI) intenta brindar una fundamentación probabilística al problema de medir constructos latentes (no observables) y considera al ítem como unidad básica de medición. La puntuación de una prueba en el modelo clásico estima el nivel de un atributo (aptitud, rasgo de personalidad, interés, etc.) como la sumatoria de respuestas a ítem individuales, mientras que la TRI se centra exclusivamente en el ítem.

Tanto la teoría clásica de los test (TCT) como la TRI abordan el mismo problema: tratar de calcular el error cometido al medir variables de naturaleza psicológica o pedagógica. Ambas teorías consideran que cada individuo lleva asociado un parámetro individual, que en la teoría de la respuesta al ítem se denomina aptitud y se simboliza por la letra griega θ (zeta), y en la teoría clásica se denomina puntaje verdadero (V), que es inobservable.

La diferencia principal entre la TCT y los diversos modelos del rasgo latente o de teoría de la respuesta al ítem, es que la relación entre el valor esperado y el rasgo en la TCT es de tipo lineal (X= V+e) mientras que en los diversos modelos de la TRI las relaciones pueden ser funciones de tipo exponencial, tales como los modelos de Poisson, de la ojiva normal, del error binomial o los modelos logísticos de 1,2 o 3 parámetros.

En la TCT las características métricas del test (tales como confiabilidad y validez) se definen en términos de un grupo determinado de personas con los que se ha construido el baremo o normas de interpretación de las puntuaciones (grupo normativo). Resulta, por tanto, muy dificultoso comparar

[1] Gil Pascual, J.A. (2008). 1ª Reemp. *Métodos de Investigación en Educación* (volumen III). Análisis Multivariante. Madrid: UNED.

individuos sometidos a distintos tests. Este problema se resuelve en la TRI cuyos resultados son independientes del grupo examinado.

3.1. Supuestos

Aunque existe modelos TRI de otra naturaleza, los TRI que haremos referencia en el texto asumen fundamentalmente dos supuestos: unidimensionalidad e independencia local. El primer supuesto exige que la respuesta al ítem esté determinada exclusivamente por una variable, denominada rasgo latente (modelos unidimensionales). La independencia local exige que la probabilidad de responder correctamente a un ítem sea independiente de la probabilidad de responder a otro cualquiera.

Para comprobar estos supuestos, sólo se mira la unidimensionalidad ya que el cumplimiento de la misma implica la independencia local.

3.2. Modelos

Los postulados básicos de la TRI son los siguientes:

1. El resultado de un individuo en un ítem puede ser explicado por un conjunto de factores llamados rasgos latentes o aptitudes que se simbolizan por θ.
2. La relación entre la respuesta de un sujeto a un ítem y el rasgo latente subyacente puede describirse como una función monótona creciente que se llama función característica del ítem o curva característica del ítem (CCI). Se cumple que a medida que la aptitud (θ) aumenta la probabilidad de una respuesta correcta al ítem también aumenta.
3. Las estimaciones de la aptitud (θ) obtenidas con distintos ítem serían iguales y las estimaciones de los parámetros de los ítems obtenidos en distintas muestras de individuos serán iguales.

Existen diversos modelos de TRI, sin embargo nos centraremos en los más utilizados los modelos donde la CCI obedece a una función logística. El más difundido es el modelo de Rasch o de un solo parámetro (b) asociado a la dificultad del ítem.

$$P_i(\theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}}$$

Donde:

$P_i(\theta)$ representa la probabilidad de acertar el ítem i para un valor de θ

b_i expresa el índice de dificultad del ítem i

D es una constante que cuando toma el valor de 1,7 la función logística se aproxima a la normal acumulada.

El parámetro θ es el rasgo latente no observado, por construcción es una variable continua cuyo rango teórico va desde – infinito a + infinito, no obstante se considera estandarizado con media 0 y desviación 1, por tanto su recorrido irá fundamentalmente entre –3 y 3.

El parámetro b es la dificultad del ítem y viene expresado en la misma escala que θ y toma el valor de θ cuando la probabilidad es (1+c)/2 donde c es el parámetro de pseudoazar del ítem. En el modelo de un solo parámetro c=0. Cuanto más elevado sea el valor de b más difícil será el ítem.

El modelo de **dos parámetros** toma la expresión:

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

En este caso además de los parámetros anteriores hay uno nuevo, a, que representa el índice de discriminación del ítem. A mayor valor de a más poder discriminativo tendrá el ítem. Su valor es proporcional a la pendiente de la CCI en el punto θ=b.

El modelo de **tres parámetros** toma la expresión:

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

El parámetro c, nuevo en este modelo, representa la probabilidad de acertar el ítem las personas sin conocimiento alguno del tema. Matemáticamente representa la probabilidad de acertar el ítem cuando $\theta \to -\infty$.

Ejemplo 2.5

Se supone un modelo TRI de Rasch donde. ¿Cuál es la probabilidad de que los estudiantes de un instituto acierten un ítem cuyo índice de dificultad b=2? (Se supone D=1).

$$P(3) = \frac{e^{1(3-2)}}{1+e^{1(3-2)}} = 0,7310$$

Aunque sólo tenemos el dato de b=2 para θ=3, suponiendo la mismas condiciones, podemos dar la curva característica del ítem:

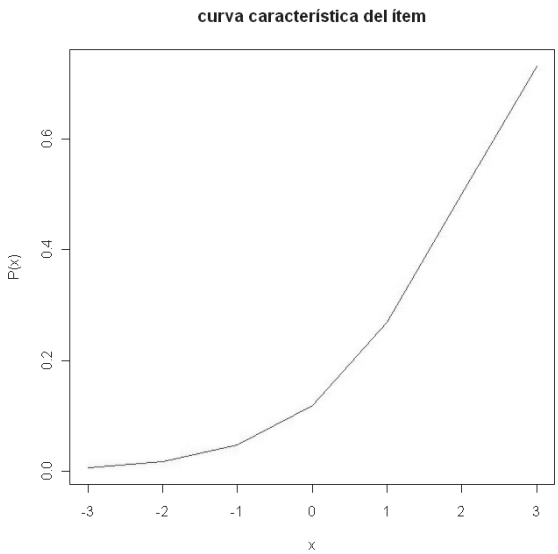


Figura 2.1. Curva característica del ítem (TRI de un parámetro).

Ejemplo 2.6

Se supone un modelo TRI de tres parámetros, se pide: la probabilidad de acertar un ítem para sujetos con θ=3, sabiendo que la probabilidad de acertar por azar es 0,3, su índice de dificultad 0,6 y su índice de discriminación es 1,3 (suponer D=1).

Naturalmente será necesario aplicar el modelo TRI de tres parámetros con c=0,3; b=0,6 y a=1,3

$$P(3) = 0,3 + (1-0,3)\frac{e^{1*1,3*(3-0,6)}}{1+e^{1*1,3*(3-0,6)}} = 0,9703$$

La curva característica del ítem será:

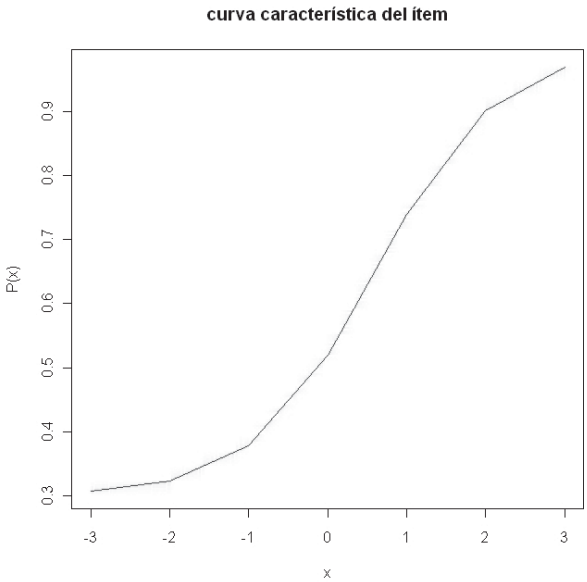


Figura 2.2. Curva característica del ítem (TRI de tres parámetros).

Desde la TRI resulta complicada la estimación de los parámetros de los modelos, dicho proceso se denomina **calibración**. Es evidente la necesaria ayuda de los programas de cómputo. Actualmente existen varios programas tales como LOGIST y BILOG-MG3, entre otros; además el programa R tiene comandos al respecto. Por último, es necesario puntualizar que para emplear modelos TRI se requieren muestras grandes (n>300) que hacen posible el ajuste a cualquier modelo de uno, dos o tres parámetros. Para muestras más pequeñas el mejor modelo es el de Rasch y de ahí su popularidad.

Para realizar el proceso de calibración, en primer lugar se tendrá que estimar los parámetros y en segundo lugar valorar la precisión de las estimaciones mediante distintos procedimientos: a) correlación entre los valores

simulados de los parámetros y los estimados; b) calcular el índice de ajuste y c) calcular la bondad de ajuste de las curvas características. Si se utiliza este último criterio se puede calcular los residuos estandarizados (RE):

$$RE = \frac{P(\theta_j) - P_e(\theta_j)}{\sqrt{P(\theta_j)Q(\theta_j)/n_j}}$$

Donde:

n_j número de sujetos en la categoría j

$P(\theta_j)$ valor de la curva característica del ítem (CCI) para el nivel θ_j

$P_e(\theta_j)$ proporción de sujetos dentro de la categoría j que superan el ítem

$Q(\theta_j) = 1 - P(\theta_j)$

Conforme RE aumente peor será el ajuste en las diferentes categorías en que se divida θ. Para ver el grado de ajuste, normalmente se fija un nivel máximo (por ejemplo RE=1,96) y se estudia el número de residuos que supera dicho nivel.

Ejemplo 2.7

En la tabla adjunta se muestran los resultados de 60 escolares en un ítem de una prueba de comprensión lectora. Los resultados se han agrupado en 6 intervalos según su puntuación en la variable latente θ. Los parámetros estimados para un modelo logístico de tres parámetros fueron los siguientes: a=0,5 ; b=0,6; c=0,1 (recordar que D=1,7)

θ	escolares
-3- -2	1 0 0 1 0 0 1 0 0 0
-2- -1	0 1 0 1 1 0 0 0 1 0
-1- 0	1 0 1 1 0 1 0 1 0 0
0- 1	1 1 1 0 0 0 1 1 1 0
1- 2	1 1 1 1 1 0 1 1 0 0
2- 3	1 1 1 1 0 1 1 1 1 1

Se toma como criterio de ajuste que ningún residuo estandarizado supere en valor absoluto a 1,96 ¿responder si el ítem de la tabla anterior se ajusta a los datos?

Para calcular RE en primer lugar se obtiene $P(\theta_j)$ tomando como valor de θ la mitad del intervalo. Si se realiza el proceso con R (recordar que en R la coma decimal se expresa con un punto) tenemos:

```
a=0.5
b=0.6
c=0.1
D=1.7
x=c(-2.5,-1.5,-0.5,0.5,1.5,2.5)
p<-c+(1-c)*exp(D*a*(x-b))/(1+exp(D*a*(x-b)))
```

Dando como resultado:

[1] 0.2358302 0.2920534 0.4120219 0.6000000 0.7879781 0.9079466

Y a continuación se realizan los cálculos posteriores:

```
q<-(1-p)
pe<-c(0.3,0.4,0.5,0.6,0.7,0.9)# proporción de sujetos que #
#superan la prueba en cada uno de los seis intervalos
re<-(p-pe)/sqrt(p*q/10)
```

Dando como resultado de RE:

[1] -0.478 -0.751 -0.565 0.000 0.681 0.087

Cuyos valores absolutos no superan el 1,96. Por tanto el modelo se ajusta a los datos.

3.3. Curva característica del test

A la hora de dar resultados de la puntuación de un individuo en una prueba, se necesita expresar la misma en función del número de ítems del test. La curva característica del test (CCT) permite transformar los valores θ en puntuaciones verdaderas:

$$PV_j = \sum_{i=1}^{n} P_i(\theta_j)$$

Donde:

PV_j representa la puntuación verdadera que corresponde a individuos con un nivel en el rasgo latente de θ_j

n el número de ítems del test

$P_i(\theta_j)$ el valor de cada CCI para $\theta = \theta_j$

Ejemplo 2.8

Para abreviar cálculos se supone un test compuesto por cuatro ítems estimados mediante R con la librería ltm, los resultados para un modelo de dos parámetros son:

a_1=1,2;a_2=1,7;a_3=2,2;a_4=2,7; b_1=1;b_2=1,5;b_3=2;b_4=2,5. Se pide calcular la CCT y la puntuación verdadera. Se supone los siguientes valores de θ=(-3,-2,-1,0,1,2,3)

Para resolver el problema se calcula $P(\theta_j)$. La tabla siguiente muestra los valores:

θ	$P(q_j)$				$PV_j = \sum_{i=1}^{n} P_i(\theta_j)$
	Ítem1	Ítem2	Ítem3	Ítem4	CCT
-3	0,000	0,000	0,000	0,000	0,000
-2	0,002	0,000	0,000	0,000	0,002
-1	0,017	0,001	0,000	0,000	0,018
0	0,115	0,013	0,001	0,000	0,129
1	0,500	0,191	0,023	0,001	0,715
2	0,885	0,809	0,500	0,092	2,286
3	0,983	0,987	0,977	0,908	3,855

Cuyo gráfico se muestra a continuación:

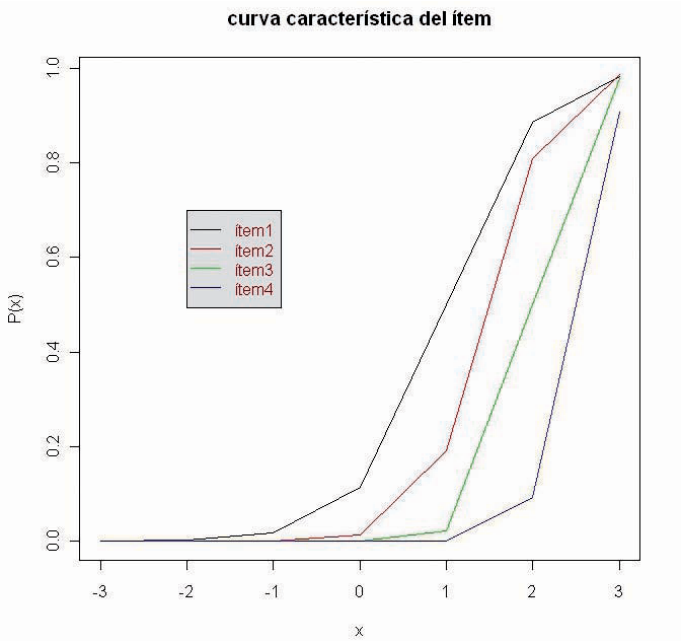


Figura 2.3. Curvas características de los ítems.

Por tanto la CCT es la relación entre la puntuación verdadera y la escala de aptitud, de ahí que si conocemos θ automáticamente sabremos calcular el valor de la puntuación verdadera para ese nivel de aptitud.

En los test de referencia criterial basados en estándares se fija la proporción de aciertos para un nivel de aptitud, que representa el grado de habilidad o conocimientos que los sujetos poseen en un dominio determinado. Su expresión es la siguiente:

$$\pi_j = \frac{1}{n}\sum_{i=1}^{n} P_i(\theta_j) \text{ en el ejemplo anterior para } \theta = 0 \text{ su valor será } 0{,}129/4$$

3.4. Funciones de información

Atendiendo a Fisher se define la información como el recíproco de la precisión en la estimación de un parámetro. En el caso de la TRI nuestro foco de atención es la estimación de θ, pero naturalmente se cometerá un error de medida (e) que será: $e = \hat{\theta} - \theta$ cuya precisión se puede expresar por la varian-

za del error de media, de ahí que la información suministrada por un ítem sobre el parámetro θ será:

$$I(\theta) = \frac{1}{\mathrm{var}(\hat{\theta} / \theta)}$$

En el modelo de un parámetro la función de información de un ítem tomará la expresión:

$$I_i(\theta) = D^2 P_i(\theta) Q_i(\theta)$$

Donde:

$I_i(\theta)$ cantidad de información del ítem i en el nivel θ

D constante de escala: 1,7

$P_i(\theta)$ probabilidad de acierto en el ítem i

$Q_i(\theta)=1- P_i(\theta)$

En el modelo de dos parámetros toma la forma:

$$I_i(\theta) = D^2 a_i^2 P_i(\theta) Q_i(\theta)$$

donde a_i es el índice de discriminación

En el modelo de tres parámetros será:

$$I_i(\theta) = \frac{D^2 a_i^2 Q_i(\theta)\left[P_i(\theta) - c_i\right]^2}{P_i(\theta)(1-c_i)^2}$$

donde c_i representa el índice de pseudoazar.

Ejemplo 2.9

En un modelo logístico de tres parámetros, para un ítem se estimaron los parámetros obteniendo los siguientes valores: a=0,7; b=1 y c=0,3. Para $\theta = 1$, ¿qué cantidad de información aporta este ítem a la medición?

Según la fórmula anterior, la cantidad de información es:

$$I_i(\theta) = \frac{D^2 a_i^2 Q_i(\theta)\left[P_i(\theta) - c_i\right]^2}{P_i(\theta)(1-c_i)^2}$$

Realizando el proceso con R tenemos:

```
a=0.7
b=1
c=0.3
D=1.7
x=1
p<-c+(1-c)*exp(D*a*(x-b))/(1+exp(D*a*(x-b)))
q=1-p
i<-(D^2*a^2*q*(p-c)^2)/(p*(1-c)^2)
i
```

cuyo resultado es:

[1] 0.1906

En los modelos de uno y dos parámetros la información será máxima para θ=b.

La **función de información de un test** será el sumatorio de las funciones de información de cada ítem del mismo:

$$IT(\theta) = \sum_{i=1}^{n} I_i(\theta)$$

Las funciones de información de un test es invariante ante el cambio de escala de θ por tanto se podrá comparar dos test mediante la llamada **eficacia relativa** (ER):

$$ER(\theta) = \frac{IT(\theta_x)}{IT(\theta_y)}$$

Donde:

$IT(\theta_x)$ información del test X para el valor

$IT(\theta_y)$ información del test Y para el valor

La función de información es un indicador de la precisión de un test, cuanto mayor sea $I(\theta)$ menor será el error típico de medida, luego será mayor la información que las estimaciones aportan al parámetro θ.

Ejemplo 2.10

Un determinado ítem de un test se ajusta a un modelo logístico de tres parámetros con a=1,2 ;b=1 y c=0,30 ¿Calcular el valor de θ que aporta más información a la medición realizada con el ítem?¿ De la respuesta a la pregunta anterior calcular la información aportada para ese θ?

Si la expresión de la función de información de tres parámetros se maximiza para θ se encuentra que el valor obtenido es:

$$\theta = b + (1/Da)\left\{\ln\left[1/2 + (1/2)\sqrt{(1+8c)}\right]\right\},$$

luego sustituyendo se tiene

$$\theta = 1 + (1/1,7*1,2)\left\{\ln\left[1/2 + (1/2)\sqrt{(1+8*0,3)}\right]\right\} = 1,172565$$

En el punto de máxima función de información la cantidad de la misma viene dada por:

$I(\theta) = [D^2a^2/8(1-c)^2][1 - 20c - 8c^2 + (1 + 8c)^{3/2}]$, luego sustituyendo

$I(\theta) = [1,7^2*1,2^2/8(1-0,3)^3][1 - 20*0,3 - 8*0,32 + (1 + 8*0,3)^{3/2}] = 22,59154$

3.4.1. Ponderación óptima de los ítems

Para maximizar la información que suministran los ítems es aconsejable su ponderación, de tal forma que la información dada por el test será:

$$I(\theta, X) = \frac{\left[\sum_{i=1}^{n} w_i P_i'(\theta)\right]^2}{\sum_{i=1}^{n} w_i^2 P_i(\theta) Q_i(\theta)}$$

y el valor de la ponderación que maximiza la información será:

Para el modelo de un parámetro: $w_i = ID$; para el modelo de dos parámetros será: $w_i = Da_i$; y finalmente para el modelo de tres parámetros: $w_i = [Da_iP_i(\theta) - c_i]/[P_i(\theta)(1 - c_i)]$.

De esta forma la puntuación total de un test en el último modelo será:

$$X = \sum_{i=1}^{n} \left\{ [Da_i P_i(\theta) - c_i] / [P_i(\theta)(1 - c_i)] \right\} (x_i)$$

4. APLICACIONES DE LA TRI

Banco de ítems

Una de las aplicaciones más utilizadas de los modelos TRI es la construcción de conjuntos de ítems de propiedades métricas conocidas.

Dos son las facetas a controlar en la conformación del banco: la construcción de los ítems donde se debe respetar la unidimensionalidad de los ítems del banco y la elaboración de los tests con unas determinadas características.

Equiparación de puntuaciones

Consiste en establecer una correspondencia entre las puntuaciones de los test, que miden la misma variable y con la misma fiabilidad.

Dentro de la TRI el establecimiento de equivalencia entre los test es innecesario, otra cuestión es establecer relación entre las puntuaciones estimadas de cada test o entre las empíricas. Para este último caso el procedimiento que se sigue es utilizar unos ítems de anclaje, los mismos en cada test y proceder en uno de los test a transformar sus puntuaciones en función de los valores obtenidos en el otro.

Funcionamiento diferencial de los ítems

Se pretende estudiar si existe distinto comportamiento de los ítems en función de diferentes grupos. Para realizar tal estudio el procedimiento es utilizar la curva característica del ítem para los dos (o más) grupos analizados y ver si existe diferencia entre las mismas. Para tal fin se puede utilizar distintos métodos uno de los más utilizados es el cálculo de las diferencias de probabilidades correspondientes a ambas curvas:

$$DP = \sum_{j=1}^{n_f} \frac{P_r(\theta_j) - P_f(\theta_j)}{n_f}$$

donde:

$P_r(\theta_j)$: es la probabilidad que los individuos del grupo de referencia tienen de superar el ítem θ_j. Este valor se obtiene de la curva característica del ítem del grupo de referencia.

$P_f(\theta_j)$: es la probabilidad que los individuos del grupo focal tienen de superar el ítem θ_j. Este valor se obtiene de la curva característica del ítem del grupo de focal.

n_f: número de individuos del grupo focal.

El valor ideal de DP es 0 toda cantidad que se aleje de tal valor denotará funcionamiento diferencial.

Tests adaptativos

Un Test Adaptativo Informatizado (o TAI) presenta los ítems y recoge las respuestas de la persona utilizando un ordenador, pero el uso del ordenador no es su rasgo más característico. Lo más genuino es su capacidad para adaptarse al rendimiento de la persona que está siendo medida, de ahí el "Adaptativo" de TAI. La idea básica es presentar exclusivamente a cada persona los ítems que nos informan de su nivel de rasgo. A una persona con mucho nivel de vocabulario inglés no tiene mucho sentido pedirle la traducción de "DOG" (perro), pues es seguro que la sabrá. Análogamente, a una persona con muy bajo nivel, tampoco se le debería preguntar por la traducción de "SLED" (trineo), pues es muy poco probable que la sepa. Un TAI evitaría ambas preguntas a personas de alto y bajo nivel de vocabulario inglés, respectivamente. Una consecuencia de sólo administrar los ítems realmente útiles es que se puede conseguir la misma precisión en la medida del rasgo con muchos menos ítems.

5. UN EJEMPLO DE CÁLCULO DE LAS CARACTERÍSTICAS TÉCNICAS DE UN TEST

Sea el siguiente ejemplo: un grupo de maestros de un colegio público pasan una prueba de matemáticas de quince ítems a un colectivo de estudiantes de cuarto de primaria. Recogen los resultados en el fichero tetra1.rda. Se pide:

1. Calcular los índices de dificultad y discriminación de cada ítem. Emitir valoración al respecto.
2. Calcular el índice de fiabilidad como consistencia interna de la escala formada por los quince ítems. Valoración al respecto.
3. Dimensionalidad del conjunto de los quince ítems teniendo en cuenta la naturaleza dicotómica de las variables. Comentarios al respecto.
4. Calcular los parámetros característicos desde la TRI.

En primer lugar, una vez cargado el fichero de datos, se deben quitar los individuos que no responden todas las preguntas y el item1, este último por no tener ningún poder de discriminación, ni dificultad al no existir ninguna variación en sus respuestas.

1. Se calculan los índices de dificultad simplemente viendo la distribución de frecuencia de cada variable. Se muestra primero la sintaxis en R y después los resultados:

```
round(apply(tetra1,2,table)*100/58,0) # tabla de porcentajes
```

```
> round(apply(tetra1,2,table)*100/58,0) # tabla de porcentajes
  MAT2 MAT3 MAT4 MAT5 MAT6 MAT7 MAT8 MAT9 MAT10 MAT11 MAT12 MAT13 MAT14 MAT15
0   48   16   34   24    7   48   43   14    26    38    12    33    45    29
1   52   84   66   76   93   52   57   86    74    62    88    67    55    71
```

El ítem MAT6 es el más sencillo y los ítems MAT2 y MAT7 los de mayor complejidad (dificultad media). Se puede decir que en conjunto la dificultad es medio-baja, pues los índices van de una dificultad media a una dificultad baja, no hay ítems con dificultad alta (índices inferiores al 50%).

A continuación se calcula el índice de discriminación de cada ítem. Los resultados se presentan de forma similar al caso anterior:

```
reliability(cov(tetra1[,c("MAT2","MAT3","MAT4","MAT5","MAT6",
"MAT7","MAT8",  "MAT9","MAT10","MAT11","MAT12","MAT13","MAT14","MAT15")],
  use="complete.obs"))
> reliability(cov(tetra1[,c("MAT2","MAT3","MAT4","MAT5","MAT6","MAT7","MAT8",
+   "MAT9","MAT10","MAT11","MAT12","MAT13","MAT14","MAT15")],
+   use="complete.obs"))
Alpha reliability =  0.8245
Standardized alpha =  0.8305

Reliability deleting each item in turn:
       Alpha Std.Alpha r(item, total)
MAT2  0.8179    0.8225         0.4023
MAT3  0.8033    0.8058         0.6351
MAT4  0.8189    0.8241         0.3829
MAT5  0.8055    0.8104         0.5682
MAT6  0.8192    0.8239         0.3864
MAT7  0.8043    0.8126         0.5698
MAT8  0.7979    0.8088         0.6472
MAT9  0.8173    0.8232         0.3986
MAT10 0.8078    0.8166         0.5328
MAT11 0.8405    0.8440         0.0921
MAT12 0.8164    0.8212         0.4202
MAT13 0.8086    0.8170         0.5188
MAT14 0.8077    0.8163         0.5294
MAT15 0.8187    0.8270         0.3819
```

La correlación ítem- total marca el índice de discriminación. Van desde 0,092 a 0,647. Los ítems que mejor discriminan son MAT8 y MAT3, el que peor discrimina es MAT11.

2. Para calcular la fiabilidad de la escala, si nos fijamos en la tabla anterior en la columna del std.alpha, si el ítem se elimina, se observa que al eliminar el ítem 11 la escala sube de valor, dando 0,844. Por tanto se calcula de nuevo los valores quitando este ítem:

```
> reliability(cov(tetra1[,c("MAT2","MAT3","MAT4","MAT5","MAT6","MAT7","MAT8",
+   "MAT9","MAT10","MAT12","MAT13","MAT14","MAT15")],
+   use="complete.obs"))
Alpha reliability =  0.8405
Standardized alpha =  0.844

Reliability deleting each item in turn:
       Alpha Std.Alpha r(item, total)
MAT2  0.8334    0.8356         0.4452
MAT3  0.8220    0.8221         0.6263
MAT4  0.8354    0.8379         0.4133
MAT5  0.8236    0.8257         0.5745
```

```
MAT6  0.8367    0.8396         0.3857
MAT7  0.8225    0.8276         0.5817
MAT8  0.8182    0.8254         0.6338
MAT9  0.8337    0.8371         0.4264
MAT10 0.8257    0.8317         0.5436
MAT12 0.8330    0.8354         0.4428
MAT13 0.8280    0.8332         0.5102
MAT14 0.8276    0.8327         0.5179
MAT15 0.8406    0.8453         0.3339
```

Y en este caso no hay ningún ítem (el MAT15 está prácticamente en ese valor) que supere dicho valor, por tanto ya está formada la escala.

3. Respecto a la dimensionalidad se debe realizar un análisis factorial tomando como dato de entrada la matriz de correlación tetracórica. Dicha matriz, al tratarse de una muestra pequeña, los valores prácticamente coinciden con la matriz de correlación pearson. No obstante veamos ambos resultados:

```
>tetra.pear<-
cor(tetra1[,c("MAT2","MAT3","MAT4","MAT5","MAT6","MAT7","MAT8","MAT9",
+   "MAT10","MAT12","MAT13","MAT14","MAT15")], use="complete.obs")

> round(tetra.pear,2)
       MAT2 MAT3 MAT4 MAT5 MAT6 MAT7 MAT8 MAT9 MAT10 MAT12 MAT13 MAT14 MAT15
MAT2   1.00 0.35 0.17 0.42 0.28 0.52 0.27 0.31  0.30  0.38  0.06  0.31 -0.09
MAT3   0.35 1.00 0.29 0.76 0.45 0.35 0.30 0.38  0.29  0.57  0.31  0.38  0.14
MAT4   0.17 0.29 1.00 0.35 0.23 0.17 0.39 0.34  0.15  0.29  0.27  0.22  0.17
MAT5   0.42 0.76 0.35 1.00 0.32 0.26 0.24 0.36  0.31  0.53  0.21  0.30  0.17
MAT6   0.28 0.45 0.23 0.32 1.00 0.28 0.18 0.29  0.15  0.32  0.10  0.17  0.12
MAT7   0.52 0.35 0.17 0.26 0.28 1.00 0.41 0.31  0.45  0.28  0.43  0.31  0.29
MAT8   0.27 0.30 0.39 0.24 0.18 0.41 1.00 0.26  0.52  0.21  0.51  0.55  0.43
MAT9   0.31 0.38 0.34 0.36 0.29 0.31 0.26 1.00  0.11  0.31  0.25  0.14  0.07
MAT10  0.30 0.29 0.15 0.31 0.15 0.45 0.52 0.11  1.00  0.14  0.51  0.34  0.40
MAT12  0.38 0.57 0.29 0.53 0.32 0.28 0.21 0.31  0.14  1.00  0.19  0.20 -0.12
MAT13  0.06 0.31 0.27 0.21 0.10 0.43 0.51 0.25  0.51  0.19  1.00  0.33  0.36
MAT14  0.31 0.38 0.22 0.30 0.17 0.31 0.55 0.14  0.34  0.20  0.33  1.00  0.33
MAT15 -0.09 0.14 0.17 0.17 0.12 0.29 0.43 0.07  0.40 -0.12  0.36  0.33  1.00
> # Paquete que permite el cálculo de la matriz tetracórica
> library(polycor, warn.conflicts = TRUE)
> cor1<-hetcor(tetra1[,c("MAT2","MAT3","MAT4","MAT5","MAT6","MAT7","MAT8","MAT9",
+   "MAT10","MAT12","MAT13","MAT14","MAT15")],ML=FALSE)

> tetra.cor<-as.matrix(cor1)
> tetra.cor<-round(tetra.cor,2)
> tetra.cor
```

```
       MAT2 MAT3 MAT4 MAT5 MAT6 MAT7 MAT8 MAT9 MAT10 MAT12 MAT13 MAT14 MAT15
MAT2   1.00 0.35 0.17 0.42 0.28 0.52 0.27 0.31  0.30  0.38  0.06  0.31 -0.09
MAT3   0.35 1.00 0.29 0.76 0.45 0.35 0.30 0.38  0.29  0.57  0.31  0.38  0.14
MAT4   0.17 0.29 1.00 0.35 0.23 0.17 0.39 0.34  0.15  0.29  0.27  0.22  0.17
MAT5   0.42 0.76 0.35 1.00 0.32 0.26 0.24 0.36  0.31  0.53  0.21  0.30  0.17
MAT6   0.28 0.45 0.23 0.32 1.00 0.28 0.18 0.29  0.15  0.32  0.10  0.17  0.12
MAT7   0.52 0.35 0.17 0.26 0.28 1.00 0.41 0.31  0.45  0.28  0.43  0.31  0.29
MAT8   0.27 0.30 0.39 0.24 0.18 0.41 1.00 0.26  0.52  0.21  0.51  0.55  0.43
MAT9   0.31 0.38 0.34 0.36 0.29 0.31 0.26 1.00  0.11  0.31  0.25  0.14  0.07
MAT10  0.30 0.29 0.15 0.31 0.15 0.45 0.52 0.11  1.00  0.14  0.51  0.34  0.40
MAT12  0.38 0.57 0.29 0.53 0.32 0.28 0.21 0.31  0.14  1.00  0.19  0.20 -0.12
MAT13  0.06 0.31 0.27 0.21 0.10 0.43 0.51 0.25  0.51  0.19  1.00  0.33  0.36
MAT14  0.31 0.38 0.22 0.30 0.17 0.31 0.55 0.14  0.34  0.20  0.33  1.00  0.33
MAT15 -0.09 0.14 0.17 0.17 0.12 0.29 0.43 0.07  0.40 -0.12  0.36  0.33  1.00
```

A continuación se realiza el análisis factorial (texto de ampliación para el estudiante) por el procedimiento de ejes principales. Siguiendo con R tenemos:

```
> ######################## Análisis factorial #################

> library(rela, warn.conflicts = TRUE)

> library(psych, warn.conflicts = TRUE)

> ##############################

> ########## Análisis factorial método ejes principales #############
> tetra.pa<-factor.pa(tetra.cor, nfactors=1, residuals = FALSE, rotate = “vari-
max”,n.obs = NA,
+       scores = FALSE,SMC=TRUE, missing=FALSE,impute=“median”,min.err = 0.001,
digits = 2, max.iter = 50,symmetric=TRUE,warnings=TRUE,fm=“pa”)

>

> tetra.pa
Factor Analysis using method =  pa
Call: factor.pa(r = tetra.cor, nfactors = 1, residuals = FALSE, rotate = “vari-
max”,
    n.obs = NA, scores = FALSE, SMC = TRUE, missing = FALSE,
    impute = “median”, min.err = 0.001, digits = 2, max.iter = 50,
    symmetric = TRUE, warnings = TRUE, fm = “pa”)
       V  PA1   h2   u2
MAT2   1 0.52 0.27 0.73
MAT3   2 0.73 0.53 0.47
MAT4   3 0.46 0.21 0.79
MAT5   4 0.67 0.45 0.55
MAT6   5 0.44 0.19 0.81
MAT7   6 0.62 0.38 0.62
```

```
MAT8   7 0.64 0.41 0.59
MAT9   8 0.48 0.23 0.77
MAT10  9 0.56 0.31 0.69
MAT12 10 0.53 0.28 0.72
MAT13 11 0.54 0.29 0.71
MAT14 12 0.55 0.30 0.70
MAT15 13 0.34 0.12 0.88

                PA1
SS loadings    3.96
Proportion Var 0.30

Test of the hypothesis that 1 factor is sufficient.

The degrees of freedom for the null model are  78  and the objective function
was  5.15
The degrees of freedom for the model are 65  and the objective function was
2.23

Fit based upon off diagonal values = 0.86
Measures of factor score adequacy
                                                PA1
Correlation of scores with factors             0.93
Multiple R square of scores with factors       0.86
Minimum correlation of factor score estimates 0.73

> summary(tetra.pa)

Factor analysis with Call: factor.pa(r = tetra.cor, nfactors = 1, residuals =
FALSE, rotate = "varimax",
    n.obs = NA, scores = FALSE, SMC = TRUE, missing = FALSE,
    impute = "median", min.err = 0.001, digits = 2, max.iter = 50,
    symmetric = TRUE, warnings = TRUE, fm = "pa")

Test of the hypothesis that 1 factor is sufficient.
The degrees of freedom for the model is 65  and the objective function was  2.23
```

La proporción de varianza explicada es del 30% inferior al criterio de Carmines y Zeller sobre la unidimensionalidad, pero el ajuste (Fit based) es del 86%, que unido a la unidimensionalidad que apuntaba el alpha de Cronbach, se puede pensar en una sola dimensión.

4. Para calcular los parámetros característicos de los ítems desde la TRI en primer lugar se calculan los descriptivos:

```
tri2 <- tetra1

# install the latest version of the `ltm' package
# (if you do not already have it)
install.packages("ltm", dependencies = TRUE)
install.packages("sfsmisc")

# load `ltm' package
library(ltm)

##########################
# Descriptive Statistics #
##########################

dsc <- descript(tri2)
dsc
```

Cuyo resultado es:

```
> tri2 <- tetra1

> # install the latest version of the `ltm' package

> # (if you do not already have it)

> install.packages("ltm", dependencies = TRUE)

package 'ltm' successfully unpacked and MD5 sums checked

> install.packages("sfsmisc")

> # load `ltm' package

> library(ltm)

This is package 'ltm' version '0.9-5'

> ##########################

> # Descriptive Statistics #

> ##########################

> dsc <- descript(tri2)

> dsc

Descriptive statistics for the 'tri2' data-set

Sample:
 13 items and 58 sample units; 0 missing values

Proportions for each level of response:
           0      1  logit
MAT2  0.4828 0.5172 0.0690
MAT3  0.1552 0.8448 1.6946
MAT4  0.3448 0.6552 0.6419
```

```
MAT5  0.2414 0.7586 1.1451
MAT6  0.0690 0.9310 2.6027
MAT7  0.4828 0.5172 0.0690
MAT8  0.4310 0.5690 0.2776
MAT9  0.1379 0.8621 1.8326
MAT10 0.2586 0.7414 1.0531
MAT12 0.1207 0.8793 1.9859
MAT13 0.3276 0.6724 0.7191
MAT14 0.4483 0.5517 0.2076
MAT15 0.2931 0.7069 0.8804

Frequencies of total scores:
     0 1 2 3 4 5 6 7 8 9 10 11 12 13
Freq 0 1 1 2 3 3 3 5 3 4  6 10  6 11

Point Biserial correlation with Total Score:
      Included Excluded
MAT2    0.5576   0.4413
MAT3    0.6847   0.6209
MAT4    0.5246   0.4097
MAT5    0.6524   0.5695
MAT6    0.4462   0.3824
MAT7    0.6716   0.5767
MAT8    0.7131   0.6283
MAT9    0.5056   0.4227
MAT10   0.6282   0.5389
MAT12   0.5159   0.4390
MAT13   0.6063   0.5057
MAT14   0.6185   0.5134
MAT15   0.4497   0.3310

Cronbach's alpha:
                 value
All Items       0.8405
Excluding MAT2  0.8334
Excluding MAT3  0.8220
Excluding MAT4  0.8354
Excluding MAT5  0.8236
Excluding MAT6  0.8367
Excluding MAT7  0.8225
Excluding MAT8  0.8182
Excluding MAT9  0.8337
Excluding MAT10 0.8257
Excluding MAT12 0.8330
Excluding MAT13 0.8280
Excluding MAT14 0.8276
Excluding MAT15 0.8406
```

```
Pairwise Associations:
   Item i Item j p.value
1       1     11   0.854
2       8     13   0.683
3       1     13   0.683
4       8      9   0.667
5       5     11   0.607
6       5     13   0.576
7       5      9   0.546
8      10     13   0.470
9       8     12   0.453
10      2     13   0.440
```

Algunos de los valores obtenidos ya los teníamos anteriormente, por ejemplo el índice de dificultad (tabla Proportions for each level) o la fiabilidad de cada ítem (tabla Cronbach's alpha).

Necesitamos ahora calcular el modelo de ajuste, primero lo haremos para una discriminación igual 1 y luego para distinto de 1.

```
> # Ajuste con discriminación igual a 1

> fit.rasch1 <- rasch(tri2, constraint = cbind(ncol(tri2) + 1, 1))

> summary(fit.rasch1)

Call:
rasch(data = tri2, constraint = cbind(ncol(tri2) + 1, 1))

Model Summary:
   log.Lik      AIC      BIC
 -355.7513 737.5026 764.2884

Coefficients:
               value std.err  z.vals
Dffclt.MAT2  -0.0735  0.3289 -0.2236
Dffclt.MAT3  -2.1521  0.4238 -5.0784
Dffclt.MAT4  -0.8225  0.3427 -2.4002
Dffclt.MAT5  -1.4713  0.3717 -3.9579
Dffclt.MAT6  -3.2117  0.5697 -5.6373
Dffclt.MAT7  -0.0734  0.3289 -0.2230
Dffclt.MAT8  -0.3449  0.3316 -1.0400
Dffclt.MAT9  -2.3188  0.4407 -5.2613
Dffclt.MAT10 -1.3519  0.3651 -3.7034
Dffclt.MAT12 -2.5015  0.4616 -5.4196
Dffclt.MAT13 -0.9212  0.3460 -2.6624
Dffclt.MAT14 -0.2548  0.3304 -0.7710
Dffclt.MAT15 -1.1337  0.3544 -3.1984
Dscrmn        1.0000      NA      NA
```

```
Integration:
method: Gauss-Hermite
quadrature points: 21

Optimization:
Convergence: 0
max(|grad|): 0.022
quasi-Newton: BFGS


> # items ordenados por dificultad, y

> # probabilidad de respuesta positiva por media individual

> coef(fit.rasch1, prob = TRUE, order = TRUE)
           Dffclt Dscrmn P(x=1|z=0)
MAT6  -3.21174775      1  0.9612740
MAT12 -2.50145044      1  0.9242434
MAT9  -2.31875936      1  0.9104188
MAT3  -2.15213167      1  0.8958678
MAT5  -1.47132740      1  0.8132591
MAT10 -1.35193656      1  0.7944461
MAT15 -1.13365938      1  0.7565136
MAT13 -0.92120798      1  0.7152882
MAT4  -0.82251613      1  0.6947702
MAT8  -0.34489417      1  0.5853789
MAT14 -0.25475939      1  0.5633476
MAT2  -0.07354943      1  0.5183791
MAT7  -0.07335681      1  0.5183310
```

```
> # Ajuste por modelo de Rasch con discriminación distinto de 1

> fit.rasch2 <- rasch(tri2)

> summary(fit.rasch2)
Call:
rasch(data = tri2)

Model Summary:
   log.Lik      AIC      BIC
 -347.2438 722.4875 751.3337

Coefficients:
               value std.err  z.vals
Dffclt.MAT2  -0.0694  0.2243 -0.3095
Dffclt.MAT3  -1.4232  0.3086 -4.6113
Dffclt.MAT4  -0.5648  0.2351 -2.4024
Dffclt.MAT5  -0.9863  0.2627 -3.7540
Dffclt.MAT6  -2.0864  0.4166 -5.0083
Dffclt.MAT7  -0.0694  0.2243 -0.3095
Dffclt.MAT8  -0.2508  0.2255 -1.1124
```

```
Dffclt.MAT9  -1.5285  0.3224 -4.7407
Dffclt.MAT10 -0.9106  0.2566 -3.5493
Dffclt.MAT12 -1.6437  0.3388 -4.8511
Dffclt.MAT13 -0.6303  0.2383 -2.6453
Dffclt.MAT14 -0.1901  0.2248 -0.8457
Dffclt.MAT15 -0.7666  0.2463 -3.1131
Dscrmn        1.7813  0.2541  7.0102

Integration:
method: Gauss-Hermite
quadrature points: 21

Optimization:
Convergence: 0
max(|grad|): 0.0025
quasi-Newton: BFGS


> # items ordenados por dificultad, y

> # probabilidad de respuesta positiva por media individual

> coef(fit.rasch2, prob = TRUE, order = TRUE)
           Dffclt   Dscrmn P(x=1|z=0)
MAT6  -2.08642808 1.781277  0.9762586
MAT12 -1.64374445 1.781277  0.9492116
MAT9  -1.52854828 1.781277  0.9383569
MAT3  -1.42320853 1.781277  0.9265681
MAT5  -0.98628580 1.781277  0.8528145
MAT10 -0.91058286 1.781277  0.8350708
MAT15 -0.76662619 1.781277  0.7966641
MAT13 -0.63034890 1.781277  0.7545126
MAT4  -0.56476073 1.781277  0.7322357
MAT8  -0.25084449 1.781277  0.6098837
MAT14 -0.19008628 1.781277  0.5838495
MAT7  -0.06943449 1.781277  0.5308812
MAT2  -0.06943162 1.781277  0.5308799
```

Ahora se verá, si el segundo modelo mejora al primero (Dscrmn=1 o Dscrmn=1,7812) :

```
> # para observar si hay diferencia significativa en Likelihood Ratio Test entre
los modeloseen the two models

> anova(fit.rasch1, fit.rasch2)

 Likelihood Ratio Table
              AIC    BIC log.Lik   LRT df p.value
fit.rasch1 737.50 764.29 -355.75
fit.rasch2 722.49 751.33 -347.24 17.02  1  <0.001

> # si no hay diferencia significativa no se mejora con el ajuste
```

Ahora probamos con el modelo de dos parámetros y vemos si mejora el ajuste:

```
> ########################
> # Ajuste modelo de 2PL #
> ########################
> fit.2pl <- ltm(tri2 ~ z1)
> summary(fit.2pl)
Call:
ltm(formula = tri2 ~ z1)

Model Summary:
   log.Lik      AIC      BIC
 -339.5733 731.1467 784.7182

Coefficients:
               value   std.err  z.vals
Dffclt.MAT2  -0.0819    0.2066 -0.3962
Dffclt.MAT3  -0.7719    5.9726 -0.1292
Dffclt.MAT4  -0.6864    0.3068 -2.2377
Dffclt.MAT5  -0.7940    0.1541 -5.1534
Dffclt.MAT6  -1.7468    0.4177 -4.1819
Dffclt.MAT7  -0.0826    0.1875 -0.4404
Dffclt.MAT8  -0.2431    0.1802 -1.3496
Dffclt.MAT9  -1.4496    0.3661 -3.9594
Dffclt.MAT10 -0.9109    0.2709 -3.3629
Dffclt.MAT12 -1.3312    0.2616 -5.0881
Dffclt.MAT13 -0.6958    0.2764 -2.5175
Dffclt.MAT14 -0.2034    0.2124 -0.9578
Dffclt.MAT15 -1.0781    0.4714 -2.2870
Dscrmn.MAT2   1.8074    0.6152  2.9379
Dscrmn.MAT3  25.2098 1604.6836  0.0157
Dscrmn.MAT4   1.2112    0.4709  2.5722
Dscrmn.MAT5   3.3275    1.4103  2.3594
Dscrmn.MAT6   2.5349    1.2838  1.9745
Dscrmn.MAT7   2.2168    0.7846  2.8255
Dscrmn.MAT8   2.3119    0.7802  2.9631
Dscrmn.MAT9   1.9011    0.7792  2.4399
Dscrmn.MAT10  1.7289    0.6399  2.7018
Dscrmn.MAT12  2.8800    1.3174  2.1862
Dscrmn.MAT13  1.4359    0.5498  2.6118
Dscrmn.MAT14  1.6916    0.6007  2.8160
Dscrmn.MAT15  0.9716    0.4364  2.2262

Integration:
method: Gauss-Hermite
quadrature points: 21
```

```
Optimization:
Convergence: 0
max(|grad|): 9.4e-05
quasi-Newton: BFGS

> coef(fit.2pl, standardized = TRUE, prob = TRUE, order = TRUE)
           Dffclt     Dscrmn    std.z1 P(x=1|z=0)
MAT6  -1.74677538  2.5349012 0.9302328  0.9882014
MAT9  -1.44960766  1.9011423 0.8850336  0.9402463
MAT12 -1.33118291  2.8800326 0.9446748  0.9788316
MAT15 -1.07811668  0.9715762 0.6968402  0.7402893
MAT10 -0.91087934  1.7289192 0.8656331  0.8284720
MAT5  -0.79396190  3.3275026 0.9576876  0.9335106
MAT3  -0.77187315 25.2097509 0.9992142  1.0000000
MAT13 -0.69579249  1.4359010 0.8206071  0.7308795
MAT4  -0.68641889  1.2112009 0.7711359  0.6966490
MAT8  -0.24313315  2.3119427 0.9178222  0.6369406
MAT14 -0.20343279  1.6915935 0.8608324  0.5851923
MAT7  -0.08255035  2.2167678 0.9115434  0.5456215
MAT2  -0.08185150  1.8073657 0.8749970  0.5369166

> # para observar si hay diferencia significativa en Likelihood Ratio Test entre
los modelos

> anova(fit.rasch2,fit.2pl)

 Likelihood Ratio Table
              AIC    BIC log.Lik   LRT df p.value
fit.rasch2 722.49 751.33 -347.24
fit.2pl    731.15 784.72 -339.57 15.34 12   0.223

> # si no hay diferencia significativa no se mejora con el ajuste
```

Como se puede observar en la salida del comando anova final no hay diferencia significativa entre el modelo de un parámetro (fit.rasch2) y el modelo de dos parámetros (fit.2pl) por tanto es el modelo de un parámetro el que se ajusta a los datos.

No obstante se puede ver el grado de ajuste mediante el estudio de los residuos:

```
> residuals(fit.rasch2)
      MAT2 MAT3 MAT4 MAT5 MAT6 MAT7 MAT8 MAT9 MAT10 MAT12 MAT13 MAT14 MAT15 Obs
 [1,]    1    1    1    1    1    1    1    1     1     1     1     1     1  11
 [2,]    1    1    1    1    1    1    1    1     1     1     0     1     1   1
 [3,]    1    1    1    1    1    1    1    1     1     1     1     0     1   2
 [4,]    1    1    1    1    1    1    1    1     1     1     1     1     0   1
```

```
 [5,]    1    1    1    1    1    1    0    1    1    1    1    0    1    1
 [6,]    0    1    1    1    1    0    1    1    1    1    1    0    1    1
 [7,]    0    1    1    1    1    0    0    1    1    1    1    1    1    1
 [8,]    1    1    0    1    1    1    1    1    1    1    1    1    1    2
 [9,]    1    1    1    1    1    0    1    1    1    1    0    1    1    1
[10,]    0    1    1    1    1    0    1    1    1    1    1    1    1    2
[11,]    1    1    1    1    1    0    1    1    1    1    1    1    0    1
[12,]    0    1    1    1    1    1    1    1    1    1    1    0    1    2
[13,]    0    1    1    1    1    1    1    1    0    1    1    1    1    1
[14,]    0    1    0    1    1    0    0    1    0    1    0    0    0    1
[15,]    0    1    0    1    1    1    0    1    1    1    1    1    1    1
[16,]    1    1    0    1    1    1    0    1    1    1    1    0    1    1
[17,]    1    1    0    1    1    1    1    1    1    1    0    1    1    1
[18,]    0    1    1    1    1    0    0    1    0    1    0    0    1    1
[19,]    1    1    1    1    1    0    0    1    1    1    1    0    0    1
[20,]    0    1    1    1    1    0    0    1    0    1    0    0    0    1
[21,]    0    1    1    1    1    0    0    1    0    1    1    0    0    1
[22,]    0    0    0    0    0    0    0    0    1    0    0    0    0    1
[23,]    1    1    0    1    1    0    0    1    1    1    0    0    0    1
[24,]    0    0    0    0    1    0    0    1    0    0    0    0    1    1
[25,]    1    1    1    0    1    1    1    1    1    1    1    1    0    1
[26,]    0    1    0    0    1    1    0    1    1    1    1    0    1    1
[27,]    0    1    1    1    1    0    1    0    1    1    1    1    1    1
[28,]    0    1    1    1    1    0    0    0    1    1    0    0    1    1
[29,]    1    1    1    1    1    0    0    1    0    1    0    0    0    1
[30,]    0    1    0    0    1    0    1    1    0    1    1    0    0    1
[31,]    1    1    0    1    1    1    0    1    1    1    0    0    0    1
[32,]    1    0    0    0    1    0    0    1    0    1    0    0    0    1
[33,]    0    1    0    1    1    0    0    1    1    0    1    1    1    1
[34,]    0    0    1    0    0    0    0    1    0    1    0    0    0    1
[35,]    0    0    1    0    1    0    0    1    0    0    0    0    1    1
[36,]    0    1    0    0    1    0    0    0    0    1    0    1    1    1
[37,]    0    0    1    0    1    0    1    1    1    1    1    1    1    1
[38,]    1    1    1    1    1    1    0    1    0    1    0    1    0    1
[39,]    0    1    1    0    1    1    1    1    1    0    1    1    1    1
[40,]    0    1    0    1    0    0    1    1    1    1    1    1    1    1
[41,]    0    0    0    0    0    0    0    0    0    0    1    0    1    1
[42,]    1    1    0    1    1    0    0    0    0    1    0    1    0    1
[43,]    0    0    0    0    1    1    0    0    1    1    1    0    0    1
[44,]    0    0    0    0    1    0    1    0    1    0    0    0    1    1

              Exp      Resid
 [1,] 10.223789647  0.2427580
 [2,]  0.605646912  0.5067287
 [3,]  1.326802173  0.5844396
 [4,]  0.475111512  0.7614989
 [5,]  0.346757460  1.1093320
 [6,]  0.309681756  1.2404851
```

```
 [7,]  0.277915736  1.3697185
 [8,]  0.680704223  1.5990533
 [9,]  0.218664217  1.6708938
[10,]  0.593885468  1.8246070
[11,]  0.171535402  2.0003089
[12,]  0.479034297  2.1975366
[13,]  0.132736777  2.3804293
[14,]  0.125660515  2.4664975
[15,]  0.115008336  2.6096035
[16,]  0.092766465  2.9786782
[17,]  0.090488607  3.0235085
[18,]  0.069027272  3.5434524
[19,]  0.065716102  3.6445420
[20,]  0.064885636  3.6710512
[21,]  0.054149787  4.0646578
[22,]  0.037132041  4.9968041
[23,]  0.036914704  5.0126232
[24,]  0.033967048  5.2415899
[25,]  0.033502609  5.2803337
[26,]  0.028894340  5.7129457
[27,]  0.028541903  5.7501955
[28,]  0.022959549  6.4480867
[29,]  0.019937549  6.9409331
[30,]  0.019676640  6.9886595
[31,]  0.015733170  7.8470196
[32,]  0.015613304  7.8780426
[33,]  0.011105538  9.3838303
[34,]  0.010775449  9.5296556
[35,]  0.007909393 11.1552615
[36,]  0.007831369 11.2115752
[37,]  0.006825621 12.0213825
[38,]  0.006704613 12.1308594
[39,]  0.004540380 14.7733017
[40,]  0.004437799 14.9446104
[41,]  0.002963085 18.3163603
[42,]  0.002474741 20.0520596
[43,]  0.001988433 22.3810296
[44,]  0.001503975 25.7469661
```

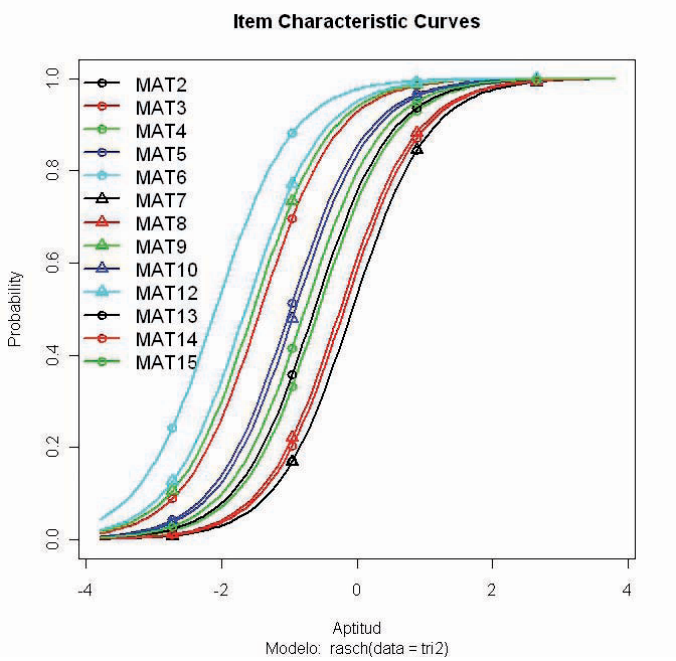
Se puede observar que las 44 posibilidades que se dan con las 13 variables (MAT2 a MAT15) de los valores 0,1, producen en algunos casos residuos notables. Por ejemplo las combinaciones binarias resultan importantes, por ejemplo, en la combinación 1,0 y en los ítems 1 y 13:

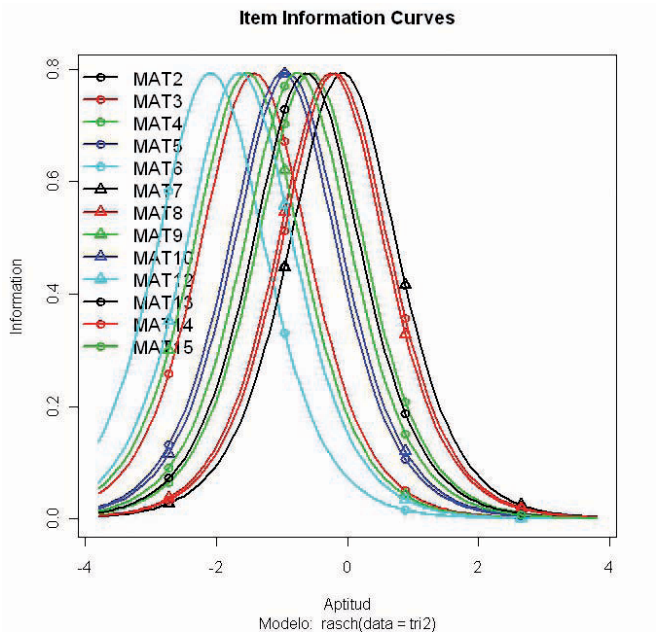
```
> # chequeo del modelo para dos marginales

> margins(fit.rasch2)

Call:
rasch(data = tri2)

Fit on the Two-Way Margins

Response: (0,0)
  Item i Item j Obs   Exp (O-E)^2/E
1      2      4   9  4.76      3.78 ***
2     10     13   1  4.37      2.60
3      1     13   7 12.37      2.33

Response: (1,0)
  Item i Item j Obs  Exp (O-E)^2/E
1      1     13  10 4.47      6.85 ***
2      1     11   9 5.25      2.68
3      2      4   5 9.10      1.85

Response: (0,1)
  Item i Item j Obs   Exp (O-E)^2/E
1     10     13   6  2.61      4.39 ***
2      2      4   0  4.18      4.18 ***
3      1     13  21 15.47      1.98

Response: (1,1)
  Item i Item j Obs   Exp (O-E)^2/E
1      1     13  20 25.69      1.26
2      1     11  21 24.91      0.61
3      2      4  44 39.96      0.41
'***' denotes a chi-squared residual greater than 3.5
```

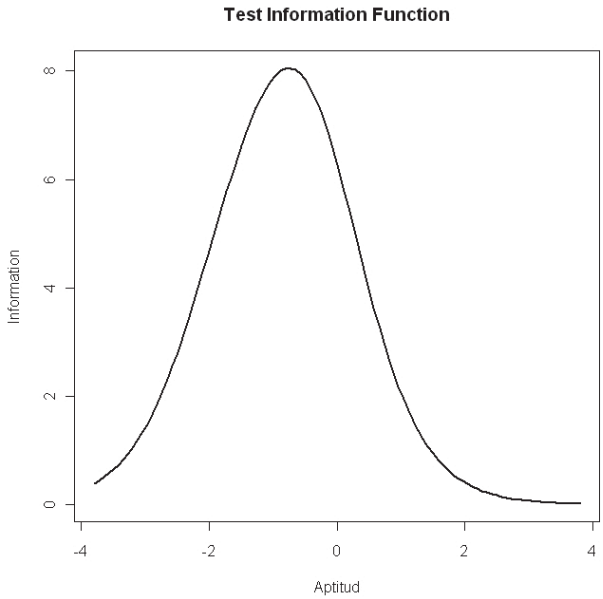
Finalmente, recordemos que los parámetros de los distintos ítems para este modelo son:

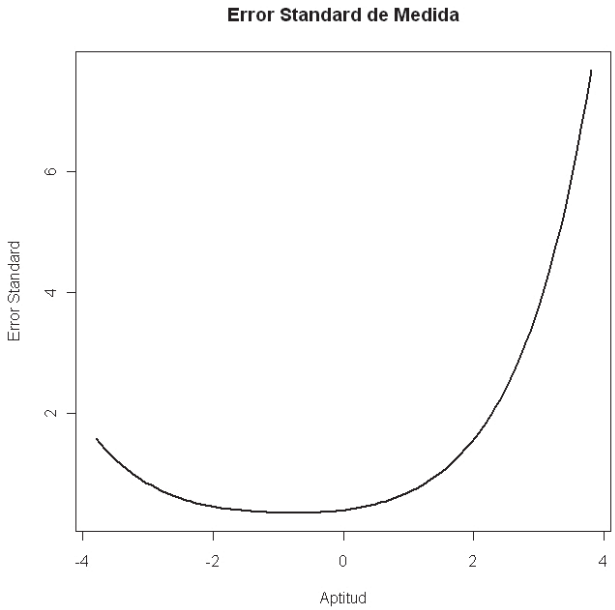
```
> fit.rasch2

Call:
rasch(data = tri2)

Coefficients:
 Dffclt.MAT2   Dffclt.MAT3   Dffclt.MAT4   Dffclt.MAT5   Dffclt.MAT6
      -0.069        -1.423        -0.565        -0.986        -2.086
 Dffclt.MAT7   Dffclt.MAT8   Dffclt.MAT9  Dffclt.MAT10  Dffclt.MAT12
      -0.069        -0.251        -1.529        -0.911        -1.644
Dffclt.MAT13  Dffclt.MAT14  Dffclt.MAT15        Dscrmn
      -0.630        -0.190        -0.767         1.781

Log.Lik: -347.244
```

Y sus curvas características son:









Claramente se puede deducir que la prueba sirve fundamentalmente para individuos de baja aptitud pero no discrimina con individuos de alta capacidad.

BIBLIOGRAFÍA

Anastasi, A. (1980). *Tests psicológicos*. Madrid: Aguilar.

Calvo, F. (1990). *Estadística aplicada: con el planteamiento y resolución de 450 problemas*. Bilbao: Deusto.

Carmines, E. G. y Zeller, R. A. (1979). *Reliability and validity assessment*. Londres: Sage.

Coombs, C. H.; Dawes, R. M. y Tversky, A. (1981). *Introducción a la psicología matemática*. Madrid: Alianza Universidad.

Cronbach, L. J. (1972). *Fundamentos de la exploración psicológica*. Madrid: Biblioteca Nueva.

De Ketele, J. M. y Roegiers, X. (1995). *Metodología para la recogida de información*. Madrid: La Muralla.

Lord, F. M. (1980). *Applications of ítem response theory to practical testing problems*. Hillsdale: LEA.

Kerlinger, F. N. (1985). *Investigación del comportamiento. Técnicas y metodología*. México: Interamericana.

Martínez Arias, R. (1995). *Psicometría: teoría de los tests psicológicos y educativos*. Madrid: Síntesis.

Muñiz, J. (2000). *Teoría clásica de los tests*. Madrid: Pirámide.

Muñiz, J. (1997). *Introducción a la Teoría de respuesta a los ítems*. Madrid: Pirámide.

Morales, P. (1988). *Medición de actitudes en psicología y educación. Construcción de escalas y problemas metodológicos*. San Sebastián: Tartalo.

Pérez Juste, R. (1986). *Pedagogía experimental. La medida en educación.* Madrid: UNED.

Siegel, S. (1970). *Estadística no paramétrica. Aplicada a las ciencias de la conducta*. Mexico: Trillas.

Capítulo 3
LA OBSERVACIÓN

1. La observación
2. El observador
3. Tipos de observación
4. Unidades de observación
5. El proceso de observar
6. Técnicas para el registro de los datos de observación
   6.1. Lista de control
   6.2. Registro anedótico
   6.3. Sistemas de rasgos
   6.4. Escalas de estimación o apreciación
   6.5. Sistemas de categorías
7. Unidades de medida
8. Análisis de validez y fiabilidad
   8.1. Fiabilidad como acuerdo entre observadores (fiabilidad inter-observadores)
   8.2. Fiablilidad entre observaciones
9. Análisis de datos
10. Un ejemplo de observación en la práctica docente

Objetivos

- Comprender la utilidad de la observación en la investigación
- Distinguir los distintos tipos de observación
- Construir instrumentos de observación
- Calcular la fiabilidad de los instrumentos de observación
- Conocer procedimientos para análisis de datos de observación

1. LA OBSERVACIÓN

Uno de los actos innatos del ser humano es la observación, de esta manera construye el conocimiento de su entorno. Forma parte del primer paso del llamado método científico aunque también se emplea en el momento de la experimentación o como técnica de recogida de datos (Buendía, Colás y Hernández, 1997).

Para algunos investigadores observación e investigación son términos disjuntos; ya que consideran la observación el paso previo que nos lleva al planteamiento de una hipótesis contrastable mediante la investigación experimental. La observación puede ser entendida como método de investigación —método observacional (Anguera, 1982)— o como técnica. En este capítulo nos centraremos fundamentalmente en la segunda acepción, es decir, en el ***proceso sistemático y controlado, mediante el cual se recoge información, en un contexto natural o ficticio, durante la realización de una investigación***. Como técnica de investigación puede ser utilizada en una investigación experimental, cuasi-experimental o expost-facto, recogiendo datos que podrán ser tratados estadísticamente.

En toda observación intervienen elementos muy concretos: el **observador**, persona que se encarga de codificar las situaciones o eventos ocurrentes; el **instrumento de registro**, que puede ser directamente los sentidos o instrumentos mecánicos como cámaras de video, fotográfica, termómetro, cinta métrica, escala de color, grabador, entre otros; y la **situación observada**, que es un complejo de múltiples eventos y relaciones, del cual hay que seleccionar los elementos a estudiar. En esta misma línea, Mario Bunge reconoce en el proceso de observación cinco elementos:

- El **objeto** de la observación
- El **sujeto** u observador
- **Las circunstancias** o el ambiente que rodean la observación

- **Los medios** de observación
- **El cuerpo de conocimientos** de que forma parte la observación

La observación se convierte en científica (Anguera,1982, p. 21) en la medida en que:

1) Sirve a un objetivo de investigación ya formulado
2) Es planificada de forma sistemática
3) Es controlada y relacionada con proposiciones más generales
4) Está sujeta a comprobaciones de validez y fiabilidad

La mayoría de los aprendizajes del **dominio cognoscitivo** son evaluados mediante procedimientos de pruebas (orales, escritas o de realización). Otras áreas, como la **comunicación** (expresión oral, expresión corporal, comunicación no verbal —gestos—, relaciones sociales), **dominio psicomotriz** (actividades deportivas, labores y destrezas manuales, artes plásticas, música y danza) y **actitudes e intereses** (actitudes sociales, científicas, culturales y afines con diferentes asignaturas, intereses, gustos y preferencias, hábitos de trabajo, etc.) se prestan a la utilización de la observación como técnica de medida.

2. EL OBSERVADOR

El problema principal en la observación es propiamente **el observador**. Debe procesar la información emanada de sus observaciones y sacar conclusiones relativas a sus construcciones hipotéticas; por tanto, puede hacer inferencias erróneas. Pero también, si el observador es objetivo y no conoce el tema de la observación, puede que lo observado no sea lo adecuado. La observación exige un conocimiento competente de lo observado y de su significado.

Otro problema es la posible influencia del observador sobre los objetos observados por el sólo hecho de que él forma parte de la situación, efecto de reactividad, (p. e. inspectores que observaban conductas de maestros pensaban que éstos trabajarían mejor en su presencia, sin embargo una persona no mejora sus aptitudes por la presencia de un observador —nadie puede hacer lo que no sabe hacer—). Las observaciones ejercen poca influencia en

las situaciones que presencian, los individuos y los grupos parecen adaptarse con rapidez a su presencia y se portan en la forma acostumbrada. El investigador ha de procurar tan sólo no entrometerse y no comunicar a las personas la sensación de ser juzgadas. Por poner un símil el observador debe ser cómo el fotógrafo de los fenómenos (Anguera, 1988, p. 7).

En general al observador se le exige el siguiente perfil:

- Conocimiento de la temática bajo estudio
- Imparcialidad
- Madurez mental, discreción e imaginación controlada
- Estar libre de toda fatiga
- Actitud alerta y activa
- Capacidad para escuchar y oír, ver y percibir
- Habilidad para pasar desapercibido, sin llamar la atención
- Comprensión de fenómenos complejos como por ejemplo el acto educativo
- Habilidades comunicativas
- Empatía
- Capacidad de autocrítica

3. TIPOS DE OBSERVACIÓN

Si atendemos a dos de los elementos relevantes de la observación —la sistematicidad o control y el grado de participación del observador—, podemos clasificar la misma en:

a) Según la sistematización o control, Anguera(1982) diferencia la observación controlada de la no controlada según se realice o no con previa sistematización o precisión científica. Según Young (s.f. en Anguera, 1982), la observación no controlada reviste de una gran importancia ya que examina cuidadosamente situaciones de la vida real, sin usar instrumentos de precisión ni comprobar los fenómenos observados. De este modo es un medio útil para la búsqueda de una problemática que oriente a una investi-

gación posterior, destacando el hecho de no influir en absoluto sobre lo observado.

— En la **observación no sistematizada, ocasional o no controlada** el observador no obedece a ninguna regla y resulta apropiada cuando el marco conceptual está poco definido por ser el primer eslabón para determinar el problema de investigación. Conforme avanza la observación se va limitando el campo además de aumentar el conocimiento y precisión de la temática objeto de estudio.

Al recoger los datos en la observación no sistematizada se debe precisar el **momento y la forma** en que debe tomar nota el observador. Se procurará que sea próximo al momento de realizar el acto de observación y en el contexto más próximo al mismo. La forma será narrativa mediante informe escrito.

— La **observación sistematizada o controlada** es la más utilizada en el ámbito de las Ciencias Humanas, parte de un estudio previo donde se han definido la situación y el problema de investigación y el observador está en condiciones de determinar las categorías aunque en el desarrollo de la investigación precise y module el número y características de las mismas.

La tipología de los datos se fija previamente. Se pueden utilizar elementos auxiliares (fotografías, registros de audio, etc.). Existe control del observador y del observado, realizando el primero la función de mediador entre la situación observada y los datos recogidos. Los datos obtenidos suelen ser cuantificables y por tanto permiten un tratamiento estadístico, y finalmente es un tipo de observación muy utilizada con grupos pequeños.

b) Conforme al grado de participación del observador, Anguera (1982) la clasifica en observación no participante, participante y auto-observación. En el primer grupo, el observador no pertenece al conjunto que se está estudiando, y puede ser directa, si se hace directamente en contacto con la realidad, o indirecta, si se basa en fuentes secundarias. La observación participante exige la presencia del investigador en el campo, y se caracteriza por la interacción social directa entre investigador e informantes, compartiendo el medio donde se desenvuelven estos últimos. Finalmente, en la auto-observación, el investigador se convierte en sujeto y objeto: observa y es observado

— La **observación interna o participante** se caracteriza por:

- Es atribuida a investigación cualitativa
- Se observa y registra "todo" lo que sucede en contacto con el contexto de observación
- Se consideran categorías de análisis emergentes
- Se actúa, registra y analiza a partir de supuestos de investigación (ontológicos, epistemológicos y metodológicos)
- Reviste mayor complejidad y énfasis en la capacidad de atención, observación y reflexión del investigador, así como los tiempos de observación y registro

— La **observación externa o no participante** se caracteriza por:

- Es atribuida a un modelo racionalista de investigación
- Se registra sólo lo preparado por un plan de observación
- Se puede observar sin contacto directo con el fenómeno o conducta. (Ej.; grabar y luego ver un video registrando). Por ejemplo al actuar en contextos familiares, conductas delictivas, etc.
- Se analiza en base a variables medibles, cuantificables definidas operacionalmente antes de observar
- Se pretende confirmar o no una hipótesis y actuar en base a un marco teórico predeterminado

4. UNIDADES DE OBSERVACIÓN

Las unidades de observación son el conjunto de eventos de conducta que se pueden adoptar a la hora de realizar un proceso de observación.

Se pueden clasificar en:

— Continuo de comportamiento

— Atributos

— Conductas

— Interacciones

— Productos de conductas

1. Continuo de comportamiento. Es la observación mediante descripción de los eventos conductuales realizada por el investigador al fijarse en una situación global, o desde distintas perspectivas. Por ejemplo:

> En una reunión de trabajo de un grupo de directivos de una Empresa Bancaria, un gerente, recientemente ascendido a director, quiere describir la situación del grupo a su entrada a la sala de reuniones. En dicha sala hay directores de sucursales que no se conocen, otros viejos conocidos que hablan y gesticulan, otros toman una taza de café, algunos juguetean con los bolígrafos a la espera del comienzo de la reunión, otros, que entran en la sala con él, ocupan unos asientos que parecían reservados para ellos. En suma, le produce una impresión global, molar, compendio de un conjunto de detalles observados.

2. Atributos. Son características deducibles de la conducta manifiesta —verbal, no-verbal o espacial. En la observación de estas unidades se suele utilizar amplios intervalos temporales. Por ejemplo: un atributo de un escolar puede ser la habilidad en la expresión hablada ("facilidad de palabra").

3. Conductas. Utilizadas como unidad de análisis de la conducta manifiesta —motora o verbal— definida en términos simples o agrupadas en clases o categorías. Se caracterizan por requerir mínima inferencia por parte del observador y por una selección rigurosa de los intervalos de tiempo, del número de sesiones y del tiempo de cada sesión. Por ejemplo: la conducta de superación observada a un escolar.

4. Interacciones. Relación secuencial entre dos eventos procedentes de dos o más personas o entre una persona y una dimensión ambiental. Se necesita la especificación previa de las conductas o clases de conductas. La observación de interacciones suele realizarse en unidades de tiempo previamente establecidas, adecuadas (temporalmente) a estas unidades de observación. Por ejemplo: la conducta de relación observada a un grupo de escolares en el patio de recreo.

5. Productos de conducta. Resultado de actividades internas o externas que los sujetos han realizado en situaciones tanto naturales como artificiales.

5. EL PROCESO DE OBSERVAR

El proceso de observación entraña un conjunto de hitos cuya secuencia es importante respetar para realizar con garantías una observación científica, dicho proceso de observación se puede formalizar mediante la siguiente secuencia:

1) Determinar el propósito de la observación:
   - Nos preguntamos si el caso o casos a investigar necesitan de la observación
   - ¿Qué unidades de observación están implicadas en el problema planteado?
   - ¿Qué unidades de medida se utilizarán en la observación?
   - Seleccionar si se precisa una muestra de las conductas a observar
2) Precisar las condiciones en que se realizará la observación
   - ¿Quién o quiénes serán observados?
     — Un solo sujeto
     — Varios sujetos
     — Controlar la representatividad del sujeto/o sujetos seleccionados
   - ¿Dónde se realizará la observación?
     — En contexto natural
     — En contexto artificial
   - Durante cuánto tiempo se realizará la observación
     — Número de sesiones
     — Duración de las mismas
     — Tiempo total de duración
   - Los observadores
     — Uno o varios observadores
     — Control del entrenamiento de los mismos

3) Precisar las técnicas de observación y análisis de datos
   - ¿Mediante qué técnicas se recogerá la información?
   - Fiabilidad y control de la generalizabilidad del proceso
   - Análisis de datos

En primer lugar, en cualquier observación se debe determinar con exactitud **el objeto a observar**, y por tanto delimitar con precisión y sin equívocos lo que se va a observar. En consecuencia será necesario definir el constructo a valorar en términos de conductas observables. Por ejemplo, si se desea medir la "asertividad" se debe determinar en qué consiste esa conducta.

Después se debe seleccionar la conducta a observar. Las conductas se pueden clasificar según la siguiente tipología:

a) **conductas no verbales**: expresiones faciales —frente y cejas, ojos, párpados y pupilas, boca y nariz—, conductas gestuales —realizados por ejemplo con la mirada— y conductas posturales —movimientos de manos, alteración del equilibrio postural, etc.

b) **conducta espacial**: colocación de las personas, distancia entre las mismas

c) **conducta extra-lingüista**: intensidad, tono y timbre de la voz; temporal —interacciones, silencios, sincronización—; continuidad —tendencia a interrumpir, dominar—; estilo verbal —pronunciación, falta de fluidez, etc.

d) **conductas lingüísticas**: palabras, frases con contenido semántico

Si se precisa se puede muestrear el tiempo, las situaciones y los sujetos a observar. Al muestrear el tiempo se debe decidir: ¿durante cuánto tiempo se realizará la observación?, ¿cuántas sesiones?, ¿en qué momentos o períodos?, ¿con qué intervalos de observación/anotación?. En las situaciones se debe responder a las siguientes preguntas: ¿cuáles son las situaciones donde se da la conducta objeto de estudio?, ¿cuántas situaciones?, ¿de qué tipo?, ¿cuándo ocurre? (por ejemplo diariamente, semanalmente). El muestreo de sujetos da respuesta: selección de intervalos en función del número de sujetos a observar, elegir un criterio de razón fija o variable.

Con estas condiciones distintos autores (Anguera, 1985, p. 72, Buendía, Colás y Hernández, 1997, p. 170) clasifican los muestreos en:

— ***Ad libitum***: "todo vale", libremente, no estructurado, no limitado (modo asistemático) generalmente utilizado al principio de una investigación, en la investigación exploratoria. No es posible determinar si las diferencias en los datos son debidas a diferencia entre los individuos o a error de medida. Por ejemplo: cuaderno de notas.

— **De eventos conductuales**: Se utilizan listas de control o escalas. Sólo se observan conductas prefijadas. No interesa el intervalo de duración de las mismas sino el número de veces que aparezca dicha conducta. Son de indudable valor cuando se pretende observar la aparición de una determinada conducta.

— **Focal**: Se define un sujeto o grupo en el que el observador centra su atención durante un intervalo de tiempo definido, registrando exhaustivamente todas las conductas objeto de interés. Aunque inicialmente se elige de forma aleatoria un sujeto se puede ir rotando entre los demás individuos del grupo.

— **De secuencias temporales**: Cuando nos interesa observar estados de conducta más amplios y perdurables. El criterio es cuantificar el tiempo de ausencia o permanencia de una conducta. Se suelen utilizar listas de control (presencia/ausencia) por la rapidez (imposibilidad de recoger más información). Puede ser instantáneo —períodos de tiempo cortos— o de intervalos. Después se necesitará fijar las condiciones en que se realizará la observación: quién, dónde, durante cuánto tiempo y quiénes realizarán la observación. También será necesario delimitar el cómo o instrumentos que se utilizarán en la observación; así como el cuándo y dónde se realizará el acto de observar.

En resumen, el proceso de observación: maneja unas unidades de análisis de la observación, se pretende un nivel de generalización, se realiza en un determinado tiempo, con un protocolo estructurado o no estructurado según el tipo de observación, por un observador u observadores entrenados, que participan o no según el tipo de observación y en una situación natural (por ejemplo en la clase) o en una situación ficticia.

6. TÉCNICAS PARA EL REGISTRO DE LOS DATOS DE OBSERVACIÓN

Los registros de observación permiten al investigador contar con un tipo de rejilla, plantilla, lista,... en la que puede apuntar, no sólo la presencia o

ausencia de determinado comportamiento, sino la intensidad y frecuencia con que se produce.

6.1. Lista de control

La lista de control es un instrumento que muestra una enumeración de una serie de características, habilidades, rasgos, conductas o secuencia de acciones cuya presencia o ausencia se desea constatar. En estas listas el investigador sólo indica si los rasgos, características, etc., están presentes o no en el objeto de estudio.

Es frecuente su empleo cuando se desea recabar información sobre: a) rasgos de comportamiento, b) resultados de una acción, de una tarea o de un aprendizaje en el que puede constatarse si determinadas características están presentes o no; y c) tareas o procesos.

6.2. Registro anecdótico

El registro anecdótico, o ficha anecdótica, es un instrumento que permite el registro de las observaciones realizadas por el investigador, expresadas de forma detallada, sobre un comportamiento generalmente poco frecuente.

En general, suelen registrarse sobre una ficha de cartulina en la que figuran en el anverso los datos de identificación (nombre, fecha, lugar... etc.), mientras que en el reverso se redacta la descripción del incidente, con toda la información que se considere relevante para el conocimiento del fenómeno objeto de estudio, así como las posibles interpretaciones del mismo.

En algunos casos no es necesario realizar un fichero con una ficha por cada anécdota y es suficiente tomar notas de forma continua pero ordenada según los incidentes de interés. Por ejemplo:

Fecha	Lugar	Incidente	Interpretación del incidente
12-11-2010 24-11-2010	Clase de 1º-A Comedor	....	....

El observador debe tener en cuenta la confidencialidad de la información, no tomar partido en el incidente y no realizar juicios de valor en base a unos pocos incidentes.

6.3. Sistemas de rasgos

Es otro sistema de observación centrado en el examen de conductas específicas, que son registradas por el observador sin emitir valoración sobre ellas. Lo que constata es la presencia o ausencia de ciertas conductas y, si lo considera oportuno, su frecuencia de aparición.

Resulta de gran utilidad para observar comportamientos poco frecuentes, al permitir un recuento de los intervalos de tiempo en que aparecen determinadas conductas.

6.4. Escalas de estimación o apreciación

Es una relación de acontecimientos concretos o categóricos, ante los cuales la respuesta se efectúa mediante un código de valoración preestablecido, que nos indica el grado de la presencia de la conducta objeto de estudio. Permite añadir datos obtenidos del propio contexto y de otros factores no analizados anteriormente (como la intensidad, duración, etc.).

Las escalas pueden ser de varios tipos:

a) Numérica.

La apreciación o estimación viene indicada por un número. Previamente se establece una equivalencia entre el valor numérico y el grado de presencia de la conducta a observar. Por ejemplo:

Se quiere valorar la satisfacción general con la materia de Estadística Aplicada. Se ha establecido las siguientes equivalencias:

1. Muy Baja Muy Pequeña
2. Baja Pequeña
3. Regular Mediana
4. Buena Grande
5. Excelente Muy Grande

b) Gráfica.

La apreciación se señala marcando una posición a lo largo de un continuo. Por ejemplo: si entrega los trabajos a tiempo

	Siempre	Casi siempre	A veces	Nunca
Entrega los trabajos a tiempo				

c) Descriptiva.

Expresa de forma clara y del modo más exacto posible la característica o rasgo que va a ser observado. Por ejemplo: se quiere observar la capacidad de trabajar en grupo y se propone cinco categorías, donde cada categoría expresa un rasgo de conducta:

	1-2	3-4	5-6	7-8	9-10
Capacidad de trabajar en grupo	Evita todo trabajo participativo o en grupo	Prefiere trabajar de forma individual	Coopera en el grupo aunque sin ningún compromiso	Denota interés cuando participa en el grupo	Siempre dispuesto a colaborar en el grupo. Subordina su éxito a la consecución de los objetivos del grupo

6.5. Sistema de categorías

El instrumento más importante de recogida de información observacional es el sistema de categorías. Un sistema de categorías es un registro sistemático de la información que exige un conocimiento profundo de la temática bajo estudio. El proceso de categorización requiere un estudio teórico con abundante trabajo de campo al estar decidiéndose qué conductas son relevantes y, en consecuencia, cuáles serán necesarias registrar para

responder a la pregunta de investigación. El proceso de categorización puede contemplar las siguientes características:

1. Las categorías deben definirse de manera clara y precisa para que su adscripción se realice con el menor error.

2. Deben ser exhaustivas. Cualquier comportamiento se puede incluir en alguna de las categorías creadas, cubriendo todo el continuum.

3. Disjuntas. Es decir, cualquier comportamiento se puede asignar a una de las categorías y no existe nexo común entre dos categorías cualesquiera.

4. Dimensionalidad. El número o los diferentes puntos de vista a tener en cuenta por los observadores.

5. Categorías continuas o discretas. Por ejemplo al clasificar la violencia de los jóvenes se pueden crear las siguientes categorías: 1) ataques verbales moderados, 2) ataques verbales y gestuales, 3) agresión física moderada, 4) agresión física violenta, etc. Estas categorías se pueden colocar en un continuo de categorías donde subyace una jerarquía de conductas definidas previamente.

6. Tamaño de la unidad. Por ejemplo unidades de tiempo. Conforme aumente el tamaño más probable son los errores del observador al asignar una conducta observada.

Conforme antes se ha señalado, las categorías se pueden definir según criterios que permiten dividir el problema en sus elementos más sustantivos, división "molecular", o atendiendo a criterios más "molares", más globales, que exige un mayor grado de conocimiento ("inferencia") del observador acerca del problema bajo estudio.

Para el tratamiento informático de datos, a las categorías obtenidas de la observación suelen asignarse numerales que permiten un tratamiento estadístico de los datos. Este proceso se suele denominar **codificación**. La codificación debe conservar las propiedades de dimensionalidad de las categorías y el orden presente.

7. UNIDADES DE MEDIDA

A continuación se presentan un conjunto de posibles medidas conductuales que se pueden extraer de un registro observacional.

a) Ocurrencia. Si el fenómeno se da o no.

b) Frecuencia. Cuántas veces se da el fenómeno o la categoría, f(j) (j=1,2,3, …n). Se pueden dar:

   1. Tasa o razón de ocurrencia. Es el número de ocurrencias por unidades de tiempo consideradas v(j)=f(j)/T.

   2. La frecuencia relativa $p(j) = \frac{f(j)}{\sum f(j)}$ es el cociente de la frecuencia de la categoría por el total de eventos registrados durante el intervalo de observación.

c) Orden. Es la sucesión de códigos relativos a las sucesivas manifestaciones de conductas. Para que exista un correcto registro es necesario que se recoja información de todo el continuo y exclusividad entre las manifestaciones de conducta, es decir, que no se produzcan los rasgos de forma simultánea.

d) Duración. Tiempo que dura el fenómeno. $D(j) = \sum d(j)$ indica el número total de tiempo que ocupan todas las ocurrencias de las categorías durante el período de observación. Para poder comparar la duración entre investigaciones, se calcula la duración relativa o prevalencia, como el cociente entre la duración de una categoría por el tiempo total de observación $\pi(j) = D(j)/T$.

e) Latencia. Tiempo entre el estímulo y la respuesta.

d) Intensidad. "Fuerza" del fenómeno. Requiere una medida ordinal que refleje los distintos grados de presencia de una determinada conducta en un sujeto.

8. ANÁLISIS DE VALIDEZ Y FIABILIDAD

El análisis de validez, en general, se construye mediante la relación establecida entre los datos recogidos de la observación y su relación con un criterio externo representativo de las conductas observadas.

La fiabilidad de los datos recogidos en los procesos de observación, se puede recoger por una parte en la estabilidad de los resultados de los sucesos observados cuando se repite la observación por otro investigador (fiabilidad test-retest) y por la estabilidad temporal de las observaciones.

8.1. Fiabilidad como acuerdo entre observadores (fiabilidad inter-observadores)

El acuerdo entre observadores viene medido por el coeficiente Kappa, variación del coeficiente de Scott (Croll, P, 1995:180):

$$\pi = \frac{P_o - P_e}{1 - P_e}$$

donde P_o es la proporción de ocasiones en las que los observadores están de acuerdo,

$$P_0 = \frac{1}{N}\sum_{i=1}^{c} n_{ii}$$

y P_e es la proporción de ocasiones en las que éste se habría dado de forma fortuita P_e es dado por:

$$P_e = \frac{1}{N^2}\sum_{i=1}^{c} n_{i.} n_{.i}.$$

Donde $n_{i.}$ y $n_{.i}$ son las observaciones marginales por cada categoría.

Por ejemplo supongamos la siguiente tabla de contingencia:

	c1	c2	c3	c4	$n_{i.}$
c1	0	1	2	0	3
c2	4	2	3	1	10
c3	2	0	4	2	8
c4	1	2	0	3	6
$n_{.i}$	7	5	9	6	27

$$P_0 = \frac{1}{27}(0+2+4+3) = \frac{9}{27} = 0,33;\ P_e = \frac{1}{27^2}(3*7+5*10+8*9+6*6) = 0,24$$

$$\pi = \frac{0,33-0,24}{1-0,24} = 0,12$$

Landis y Koch propusieron unos márgenes para valorar el grado de acuerdo en función del índice de kappa:

kappa	**Grado de acuerdo**
< 0	sin acuerdo
0-0,2	insignificante
0,2-0,4	bajo
0,4-0,6	moderado
0,6-0,8	bueno
0,8-1	muy bueno

De acuerdo con la tabla anterior el porcentaje de acuerdo es insignificante.

8.2. Fiabilidad entre observaciones

Una forma de controlar la estabilidad en las observaciones será mediante las correlaciones que puedan obtenerse entre dos observaciones registradas en dos o más momentos mediante el coeficiente de correlación de Pearson o Spearman (test-retest).

9. ANÁLISIS DE DATOS

En el análisis de datos será necesario recoger la información de las unidades de medida con lo que se pueden realizar una gran variedad de análisis cuantitativos. El tipo de análisis a utilizar depende en primera instancia de los objetivos e hipótesis de investigación planteadas y en segundo lugar de las posibilidades informativas de los datos registrados (sólo frecuencia, duración, secuencia de categorías, ...).

Existen distintas clasificaciones de las tipologías de análisis de datos a realizar. La mayoría de autores, con matizaciones, recogen la propuesta de Quera(1993):

Según la inclusión de la variable tiempo: **análisis sincrónico** *versus* **análisis diacrónico**. En el primer caso estamos hablando de medidas tomadas simultáneamente. Se describirán las relaciones de ocurrencia en un mismo punto temporal (hora, minuto, segundo, día, semana, mes, año,..). En análisis diacrónico, se relacionan medidas tomadas en diferentes puntos temporales.

Según las medidas conductuales utilizadas: **macro-análisis** *versus* **micro-análisis**. En el primer caso estamos hablando del estudio descriptivo y/o relacional de medidas globales —frecuencias, duraciones, prevalencias, …— . Si estas relaciones de medidas globales se estudian en el mismo momento temporal estamos hablando del **análisis transversal** y si el estudio es en diferentes momentos temporales estamos hablando de los **análisis longitudinales**.

Las técnicas estadísticas utilizadas para cuantificar las relaciones va desde el análisis de correlación bivariada, hasta análisis multivariantes como el factorial, correspondencias o conglomerados.

En el micro-análisis se estudian las relaciones entre las sucesivas unidades de observación reflejo del comportamiento del individuo o del grupo en el continuo del tiempo. Se pretende con este micro-análisis o en su forma extensa **análisis secuencial** ver cómo cambian las conductas en función de la ocurrencia de otras precedentes y se pretende encontrar regularidades de comportamiento, es decir, se buscan patrones estadísticos modelos de la regularidades subyacentes.

En ocasiones se utiliza el micro-análisis desde una perspectiva sincrónica, porque se desea saber la probabilidad de coincidencia temporal de varias conductas y estudiar la relación entre las mismas en un momento concreto.

Como técnicas estadísticas utilizadas en el análisis secuencial están las técnicas de retardo, las cadenas de markow o el análisis de series temporales. Si el análisis secuencial se plantea sobre una tabla de contingencia multidimensional se pueden aplicar técnicas relacionadas con los modelos log-lineales. Los paquetes estadísticos clásicos como SPSS,SAS o R tienen módulos para el tratamiento de datos secuenciales, no obstante el paquete SDIS-

GSEQ de Bakeman y Quera (1996) está específicamente preparado para tratar dichos datos.

10. UN EJEMPLO DE OBSERVACIÓN EN LA PRÁCTICA DOCENTE

En la práctica docente de enseñanza de las matemáticas durante la práctica de resolución de problemas en clase, un profesor interactúa con sus alumnos y toma nota de las categorías generadas en dichas interacciones:

1 = el profesor enuncia un problema

2 = el alumno pregunta

3 = el profesor pregunta

4 = el alumno comenta con sus compañeros

5 = el alumno responde a las preguntas del profesor

6 = el alumno trabaja

Cada cinco segundos un observador toma nota del estado o categoría en que se encuentra el proceso.

Los resultados de 40 observaciones simuladas se muestran a continuación:

3 1 4 5 5 1 3 5 2 5 3 2 5 1 2 6 2 1 6 5 2 4 3 2 1 1 6 6 6 1 6 3 4 5 4 3 1 6 1 6

Como conclusión de las observaciones se puede extraer en primer lugar la frecuencia de cada categoría:

x					
1	2	3	4	5	6
9	6	6	4	7	8

Y luego las interacciones que se dan entre dos categorías, es decir, las secuencias de categorías contiguas generan la tabla siguiente:

	1	2	3	4	5	6	
1	1	1	1	1	0	5	9
2	2	0	0	1	2	1	6
3	2	2	0	1	1	0	6
4	0	0	2	0	2	0	4
5	2	2	1	1	1	0	7
6	2	1	1	0	1	2	7
	9	6	5	4	7	8	39

Según se puede observar en la tabla, las categorías contiguas que más se dan, es la 1 seguida de la 6.

El proceso realizado para agrupaciones de dos categorías se puede generalizar para tres, cuatro, …. y naturalmente generará sus conclusiones pertinentes.

BIBLIOGRAFÍA

Anguera, M. T. (1988). *La observación en la escuela*. Barcelona: Grao.

Anguera, M. T. (1985). *Manual de prácticas de observación*. México: Trillas.

Anguera, M. T. (1982). *Metodología de la observación en las Ciencias Humanas*. Madrid: Cátedra.

Angulo Rasco, J. F. (1990). "Una propuesta de clasificación de las técnicas de recogida de información, con especial referencia a las técnicas de investigación interpretativas". En J. B. Martínez Rodríguez (ed.), *Hacia un enfoque interpretativo de la enseñanza* (pp. 85-93). Granada: Servicio de Publicaciones de la Universidad de Granada.

Arnau Grass, J. et al. (1990). *Metodología de la observación en las ciencias del comportamiento*. Murcia: Universidad de Murcia.

Bakeman, R. y Gottman, J. M. (1989). *Observación de la interacción: Introducción al análisis secuencial.* Madrid: Morata.

Bakeman, R. y Quera, V. (1995). *Con análisis de la Interacción: análisis secuencial SDIS y GSEQ.* Madrid: Rama.

Barbero, M.I. (1993). *Psicometría II. Métodos de elaboración de escalas.* Madrid: UNED.

Bisquerra, R. (1989). *Métodos de Investigación educativa. Guía práctica*. Barcelona: Ceac.

Briones, G. (1982).*Métodos y técnicas de investigación para las ciencias sociales.* México: Trillas.

Buendia, L.; Colás, P. y Hernández, F. (1997). *Métodos de investigación en Psicopedagogía.* Madrid: McGraw-Hill.

Bunge, M. (1976). *La investigación científica*. Barcelona: Ariel.

Cicourel, A. V. (1982). *El método y la medida en Sociología*. Madrid: Editora Nacional.

Coombs, C. H.; Dawes, R. M. y Tversky, A. (1981). *Introducción a la psicología matemática*. Madrid: Alianza Universidad.

Croll, P. (1995). *La observación sistemática en el aula*. Madrid: La Muralla.

Cronbach, L. J. (1972). *Fundamentos de la exploración psicológica*. Madrid: Biblioteca Nueva.

De Ketele, J. M. y Roegiers, X. (1995). *Metodología para la recogida de información*. Madrid: La Muralla.

Del Rincón, D. y otros (1995). *Técnicas de investigación en Ciencias Sociales*. Madrid: Dykinson.

Duverger, M. (1974). *Método de las ciencias sociales*. Barcelona: Ariel.

Evertson, C. M. y Green, J.L. (1986). Observation as inquiry and method. En M. C. Wittrock (ed.), *Handbook- of research on teaching* (pp. 162-213). New York: MacMillan Pub. Co.

Fernández Ballesteros, R. et al. (1992). *Introducción a la Evaluación Psicológica (2 Vols).* Madrid: Pirámide.

Fontana, A. y Frey, J. H. (1994). "Interviewing: The Art of Science". En N, K. Denzin y Y. S. Lincoln (eds.), *Handbook of Qualitative Research* (pp. 361-376). Londres: Sage.

García Cueto, E. (1993). *Introducción a la Psicometría*. Madrid: Siglo XXI de España.

García Ramos, J. M. (1986). "Validación de constructo en el ámbito Pedagógico". *Revista Española de Pedagogía*. 174, pp. 535-554.

García Ramos, J. M. (1989). *Bases pedagógicas de la evaluación. Guía práctica para educadores*. Madrid: Síntesis.

Kerlinger, F. N. (1985). *Investigación del comportamiento. Técnicas y metodología*. México: Interamericana.

López Feal, R. (1986). *Construcción de instrumentos de medida en Ciencias Conductuales y Sociales.* Barcelona: Alamex.

Martínez Arias, M. R. (1995). *Psicometría: Teoría de los Tests Psicológicos y Educativos*. Madrid: Síntesis.

Morales Vallejo, P. (1988). *Medición de actitudes en Psicología y Educación*. San Sebastián: Tártalo.

Nunnally, J. C. (1978). Psychometric Theory. Nueva York: McGraw-Hill (Traducido por Trillas 1987).

Osgood, G., Suci, G. y Tannembaum, P. (1976). *La medida del significado*. Madrid: Gredos.

Pérez Serrano, G. (1994). *Investigación cualitativa. Retos e interrogantes*. (2 vol.) Madrid, La Muralla.

Quera, V. (1993). "Análisis secuencial". En M. T. Anguera (ed.). *Metodología observacional en la investigación psicológica*, vol. 2, pp. 349-586. Barcelona: PPU.

Rincón, D. del; Arnal, J.; Latorre, A. y Sans, A. (1995). *Técnicas de investigación en Ciencias Sociales*. Madrid: Dykinson.

Sierra Bravo, R. (1985). *Técnicas de investigación social. Teoría y ejercicios*. Madrid: Paraninfo.

Capítulo 4
TEST Y PRUEBAS OBJETIVAS

1. Introducción
2. Test
3. Pruebas objetivas
4. Construcción de las pruebas
   4.1. Objetivos de los test o pruebas
   4.2. Especificación de la prueba
   4.3. Confección de la prueba
5. Aplicación de la prueba
6. Puntuación de la prueba
   6.1. Puntuaciones de los ítems
   6.2. Puntuación de la prueba
7. Interpretación de las puntuaciones de la prueba
8. Clasificación de los test

Objetivos

- Aprender el significado de los test y pruebas objetivas
- Conocer las etapas de construcción y aplicación de los test o pruebas
- Conocer estrategias para confeccionar buenos ítems
- Saber construir test y pruebas objetivas en el contexto educativo

1. INTRODUCCIÓN

Estos dos instrumentos de medida se emplean con exhaustividad en Psicología, Sociología y Educación. Las diferencias más importantes entre estos ***dos indicadores*** radican en que las pruebas objetivas están dirigidas a la medida de conocimientos, y los test a la medida de aptitudes o rendimiento.

Durante el presente capítulo se utilizará indistintamente ambos términos con la denominación genérica de prueba, sólo se distinguirá cuando expresamente se necesite.

2. TEST

Con el nombre de test se engloba todo un conjunto de instrumentos que se vienen utilizando en la investigación educativa con la intención de recoger información sobre las capacidades y/o habilidades de los sujetos. A través de los test podremos "medir" una muestra de las características o conductas de un sujeto o grupo.

Respecto a los test, Yela (1984) los define como "una situación problemática, previamente dispuesta y estudiada, a la que el sujeto ha de responder siguiendo ciertas instrucciones y de cuyas respuestas se estima, por comparación con un grupo normativo (o un criterio u objetivo), la calidad, índole o grado de algún aspecto de su personalidad". Esta definición nos lleva a destacar como características más importantes de los test:

a) Seguridad y precisión de la medida

b) Elección de una muestra de conductas, representativas del rasgo o conducta a medir

c) Comparación de conductas con un grupo normativo (Teoría clásica de los test)

d) Predicción o inferencia de conductas más generales e importantes que las observadas en el test

En investigación educativa resultan útiles para:

- Clasificar a los sujetos y grupos en diversas categorías
- Analizar diferencias intra-grupos, inter-grupos, interindividuales o intra-individuales
- Verificar hipótesis
- Predecir comportamientos
- Seleccionar individuos o grupos

La calidad de estos instrumentos, fundamentalmente por su valor predictivo, se refleja en unos requisitos técnicos de validez, fiabilidad, baremación, etc.; además del respeto escrupuloso a las normas de aplicación.

3. PRUEBAS OBJETIVAS

Cuando en las investigaciones queremos medir alguna variable para la cual no se dispone de ningún test o prueba adecuada, o cuando necesitamos evaluar el rendimiento de los estudiantes, en estos casos no queda más remedio que preparar una prueba "ad hoc", a este tipo de prueba se denomina prueba objetiva, aún dicho instrumento necesita de unos condicionantes que se expondrán a continuación.

4. CONSTRUCCIÓN DE LAS PRUEBAS

El ciclo de elaboración y aplicación de un test o prueba es el siguiente:

1. Construcción de la prueba
2. Versión previa de la prueba
3. Administración de la prueba
4. Evaluación del test o prueba
   <"Si el resultado es favorable pasar al punto 5">
   <"Si el resultado es desfavorable pasar al punto 1">
5. Versión definitiva de la prueba

La construcción de un test o prueba exige del investigador la respuesta a un conjunto de interrogantes: ¿cuáles son los objetivos de la prueba?, es decir, ¿qué se va a medir?, ¿a quién se va a medir?, ¿cuál es la utilidad de la medición?. Necesitará también especificar el contenido de la prueba, el formato de los ítems, el número de ítems, y las características métricas esperadas y finalmente se debe confeccionar los ítems.

Según Bisquerra (1989, p. 97), siguiendo el modelo psicométrico clásico, la elaboración de estas pruebas pasa por las siguientes fases:

1. Formular los objetivos de la prueba, o en su caso el constructo hipotético a medir
2. Definir la población objeto de la prueba
3. Muestreo de preguntas a partir de una tabla de especificaciones
4. Construcción de los ítems, de la forma más adecuada para lo que se quiere medir. Elaborar un banco de ítems
5. Prueba piloto sobre una muestra pequeña de sujetos
6. Primer análisis de ítems y revisión de la prueba
7. Experimentación de la prueba sobre una muestra representativa de sujetos
8. Análisis de datos: cálculo de la validez, fiabilidad, análisis de ítems y baremación
9. Elaboración del manual de instrucciones

El material impreso de la prueba constituye el denominado ***protocolo;*** en él se incluyen los ítems a resolver junto a datos de identificación.

4.1. Objetivos de los test o pruebas

Como se ha comentado el objetivo de la prueba es saber características de la variable de interés, la población objetivo o a veces también llamada diana, y la utilidad perseguida con la prueba.

La **variable de interés** puede ser una variable manifiesta, es decir, medible directamente o una variable latente —no medible directamente—. Puede ser la variable del dominio cognitivo o actitudinal.

La **población objetivo o diana** será el grupo de individuos al cual va destinada la prueba. Por ejemplo debe ser distinta la prueba de rendimiento que se pase a unos alumnos al comienzo del curso, con la prueba seleccionada para valorar el rendimiento al final del mismo. En suma, la prueba debe estar de acuerdo con el grado de maduración y características de la población objetivo.

Respecto a la **utilidad** es múltiple el abanico de posibilidades de los test o pruebas. El investigador determinará el valor objetivo de la prueba que desea construir. De forma general en la escuela los test de inteligencia y de evaluación de conocimientos son aplicados de forma rutinaria para asesorar las decisiones individuales y para mejorar la educación y organizar los planes de estudio. También en las escuelas primarias se utilizan los test audiovisuales para determinar la capacidad del alumno para aprender a leer y escribir. Si surge algún problema de desarrollo intelectual del estudiante los test ayudan al diagnóstico de los problemas. En la educación secundaria, muchos centros de enseñanza aplican test de intereses y aptitudes para orientar profesionalmente al estudiante.

En el ámbito clínico, los test sirven para proporcionan información sobre el funcionamiento y estructura de la personalidad de un paciente y la conveniencia de aplicar algún tipo de psicoterapia.

Los test también se emplean en las empresas para la selección y clasificación del personal. A los test de inteligencia o de personalidad suelen añadirse pruebas específicas, propias de las competencias exigidas por el puesto a cubrir.

4.2. Especificación de la prueba

Para especificar la estructura de la prueba se necesita tomar decisiones sobre:

1. El contenido de la prueba
2. El formato de los ítems que conformarán el test o prueba
3. El número de ítems de la prueba
4. Características métricas esperadas

Además será necesario tener en cuenta el contexto donde se aplicará la prueba: tiempo de duración de la prueba, tipo de aplicación de la misma (individual o colectiva), medio soporte de la prueba (papel, web, teléfono, etc.), etc.

En **el contenido** el investigador fijará las características o el constructo que se desea medir. La situación ideal sería conocer, por la teoría subyacente, las particularidades de dicho constructo y por tanto estaría totalmente definido el mismo, pero esta situación ideal rara vez se da, en consecuencia es necesario tener una estrategia alternativa. La herramienta más utilizada sobre todo en el dominio cognitivo es la llamada **tabla de especificación del test o prueba.**

La tabla de especificación consiste en una tabla de doble entrada donde en filas se indican las operaciones o procesos cognitivos presentes en el constructo a medir. Según la taxonomía de Bloom (1956) existe una jerarquía en dichas operaciones según el siguiente orden de menor a mayor:

1. Conocimiento
2. Comprensión
3. Aplicación
4. Análisis
5. Síntesis
6. Evaluación

En columnas las áreas de contenidos relevantes del constructo, las celdillas indicarán el peso que el investigador da a los diferentes cruces de áreas y procesos cognitivos.

Por ejemplo: se está diseñando una prueba de rendimiento en estadística aplicada. Los profesores fijan el reparto de los pesos de los ítems de la prueba según su valoración subjetiva de importancia. El resultado se muestra en la tabla 4.1.

Tabla 4.1. Tabla de especificación de la prueba de rendimiento

	Estadística descriptiva	**Probabilidad**	**Inferencia paramétrica**	**Inferencia no paramétrica**	**Análisis varianza**	**Total**
Conocimientos	3	3	2	2	2	12
Comprensión	2	1	1	1	1	6
Aplicación	1	1	2	2	2	8
Análisis	1	1	1	1	1	5
Total	7	6	6	6	6	31

El **formato de los ítems** se debe acomodar al ámbito de actuación. En el *ámbito cognitivo* se utilizarán, básicamente los siguientes tipos de ítems:

Tabla 4.2. Formato de los ítems en el ámbito cognitivo

Ítems de elección	Múltiple. Se elige la opción adecuada entre tres o más opciones	Por ejemplo: la operación de adición o suma se utiliza para: a) comparar dos cantidades b) agregar dos cantidades homogéneas c) agregar dos cantidades heterogéneas d) separar dos cantidades homogéneas
	Binaria o alternativa. Sólo hay dos posibles opciones de respuesta	Por ejemplo: la distribución normal es simétrica a) verdadero b) falso
	Emparejamiento. Se trata de enlazar los resultados de una columna con los resultados de otra columna. No tienen por qué cubrirse todos los resultados	Por ejemplo: Columna A — Columna B Media — Medidas de forma Varianza — Medidas tendencia central Desviación tip. — Medidas de dispersión Mediana — Medidas de apuntamiento Curtosis
Ítems de construcción	Ítems de respuesta corta o de completar	Por ejemplo: La media es una medida de ——— ——- obtenida mediante la suma de todos los valores de la variable dividido entre el ———.
	Ítems de respuesta extensa	Por ejemplo: defina en pocas palabras la mediana

En este ámbito será difícil elegir el ítems más apropiado para los objetivos marcados, no obstante, se suele elegir ítems de elección cuando se

intenta trabajar los dominios de conocimientos, comprensión y aplicación y se reserva los ítems de construcción para el resto de los dominios.

Si se quiere evaluar variables del *ámbito no cognitivo* (valores, personalidad, opiniones, actitudes, intereses, etc.) se suele utilizar dos tipos de ítems:

a) Ítems de respuesta abierta. Donde la persona debe escribir la respuesta deseada. Por ejemplo: valorar su opinión sobre las actividades extraescolares para el niño.

b) Ítems de respuesta cerrada. Debe contestar una opción entre varias. Dentro de los ítems de respuesta cerrada hay dos variantes:

b1) Escalas de categorías. Donde la respuesta a una pregunta se presenta en un conjunto de categorías, ordenadas o no. Cuando se están midiendo actitudes y las categoría están ordenadas según el grado de acuerdo a la pregunta planteada, tenemos las escalas sumativas o de Likert. Por ejemplo: la falta de atención en clase es el mayor problema en los institutos de secundaria:

a) Totalmente de acuerdo
b) De acuerdo
c) Ni de acuerdo, ni en desacuerdo
d) En desacuerdo
e) Totalmente en desacuerdo
f) No sabe/no contesta

Los apelativos en las escalas de categorías pueden atender: al grado de acuerdo, a la frecuencia de uso, a la cantidad, a los sentimientos o a la valoración (escala numérica y no numérica).

b2) Listados. Es un conjunto de frases relativas a un hecho, sujeto o idea donde se debe marcar las respuestas más acordes con la opinión del sujeto respecto al hecho o idea a valorar. Por ejemplo: en su opinión marque los problemas más importantes con que se enfrenta en el trabajo:

a) Falta de motivación
b) Baja remuneración
c) Alto nivel de exigencia

d) Falta de compatibilidad con la vida familiar

e) Excesiva distancia de su domicilio

f) Pocos permisos

g) Trato inapropiado por la jefatura

Una variante de los listados se tiene cuando se pide ordenar la lista de las respuestas propuestas según el grado de importancia para el individuo.

Respecto a la **longitud de la prueba,** se debe tener presente, en primer lugar, la tabla de especificaciones donde se ha fijado los pesos de cada celda. Una primera aproximación será, por ejemplo, dar un número de ítems por celdilla. Si por ejemplo, se fija en 2 ítems por celda, el ejemplo de la tabla 4.1 tendrá 62 ítems. Será necesario valorar la duración de la prueba para evaluar la viabilidad de la misma, suponiendo unos valores estándar de 1 minuto, si el ítems es de elección múltiple, y de 3 minutos respuesta corta o elección binaria, para un adolescente o adulto.

También la longitud de la prueba está influenciada por la fiabilidad deseada para la misma, ya que, según sabemos, a mayor longitud mayor fiabilidad, en consecuencia esos dos factores nos harán dimensionar la longitud de la prueba.

Las **características métricas deseadas** marcarán finalmente el reparto de los ítems porque si la prueba está orientada para un nivel de conocimientos se debe saber el nivel de dificultad de los ítems y la discriminación de los mismos. En el capítulo 2 del texto se ha tratado en detalle estas características. En este caso sólo será necesario determinar, por ejemplo, si la prueba se desea orientar para un nivel general de conocimiento, será necesario eliminar los ítems con dificultad alta o baja.

4.3. Confección de la prueba

4.3.1. Confección de los ítems

La construcción de los ítems se debe realizar por personas expertas en la temática y conocedoras de la población diana. Se aconseja crear unas tres veces más de ítems que los necesarios y cuando menos el 25% más.

A la hora de confeccionar los ítems hay unas reglas generales que se incluyen en cualquier manual dedicado a esta temática, alguna de las cuales exponemos a continuación:

- El lenguaje utilizado debe ser claro, sencillo y redactado de tal forma que pueda comprenderse con facilidad (no utilizar términos técnicos).
- Se debe procurar evitar negaciones en el enunciado de los ítems.
- Si se quieren respuestas concretas las preguntas deben ser concretas.
- Expresar una única idea por ítem.
- No deben incorporar términos morales (juicios de valor).
- Tampoco deben formularse preguntas de doble efecto, es decir preguntas cuya redacción pueda implicar una doble respuesta.

En el caso de utilizarse la prueba en el ámbito cognitivo, además de las recomendaciones generales será necesario tener en cuenta alguna de las siguientes premisas:

- Hay que evitar como respuesta la opción ninguna de las anteriores o todas las anteriores.
- Controlar la posición en que aparece la opción correcta para que no salga de forma sistemática en el mismo lugar.
- Evitar dar pistas sobre la opción correcta.

En el ámbito no cognitivo, alguna de las recomendaciones se verán en el capítulo 5 dedicado a las encuestas y en concreto a los ítems de las mismas. Se deberá controlar los sesgos en las respuestas de los sujetos: sesgos relativos a la deseabilidad social —responder aquello que tenga mayor grado de aceptación social—, la aquiescencia —responder sistemáticamente siempre de acuerdo o siempre en desacuerdo— y finalmente la respuesta extrema —responder siempre alguna de las respuestas extremas de forma independiente del ítem respondido—.

4.3.2. Confección de la prueba

En la confección final de la prueba se debe tener en cuenta: las instrucciones para responder la prueba, el formato de presentación de la misma y el elemento donde se registrarán las respuestas de la prueba.

Las **instrucciones de aplicación** de la prueba tendrán como objetivo la comprensión de la mecánica de la misma para que no se produzca ningún error por desconocimiento de ésta y no de su contenido. Si se necesita vendrá ilustrada con algún ejemplo aclaratorio y tendrá distinta redacción según se realice la prueba por correo o web y de forma presencial. Normalmente si la prueba se realiza por correo o web será necesario incluir mayor cantidad de información para facilitar su cumplimentación sin equívocos.

Tabla 4.3. Ejemplo de instrucciones de aplicación de una prueba

1. INSTRUCCIONES GENERALES

El local donde se aplique la prueba debe reunir las condiciones suficientes de amplitud, iluminación y ventilación. Los estudiantes se colocarán de forma que puedan trabajar con independencia. La prueba es aconsejable realizarla a primera hora de la mañana, es decir, sobre las 9,00 h.

La valoración de la prueba (utilizando la plantilla preparada para este menester), se llevará a cabo en otro momento fuera de la situación de aplicación del instrumento de medida.

Es necesario que el examinador tenga preparado, de antemano, todo el material necesario, cumplimentando los datos de identificación y, por supuesto, que esté perfectamente entrenado en la aplicación y corrección para que dirija la prueba sin interrupciones ni titubeos. Asimismo debe crear en el alumno un clima de confianza para que trabaje con tranquilidad y seguridad deslindando que los resultados de la prueba no influirán en las calificaciones y que por supuesto son secretos.

2. INSTRUCCIONES ESPECÍFICAS DE APLICACION

El día antes de la aplicación el profesor debe recordar a los estudiantes que al día siguiente traigan lapiceros y goma de borrar. Con tiempo suficiente debe tener previsto todo el material necesario para la aplicación: los cuadernillos, las hojas de respuesta, un reloj cronómetro y lapiceros y gomas por si alguno se le olvida los suyos.

Se indicará al estudiante que en la hoja de respuestas debe cumplimentar los datos de identificación que conozca, y si tiene alguna duda estos serán rellenados por el profesor.

Una vez que los alumnos estén en sus asientos, con el material anteriormente indicado, se les pedirá que a partir de ese momento guarden silencio y que si alguno tiene que preguntar algo que levante la mano para que el profesor le aclare sus dudas personalmente

El profesor dirá a continuación lo siguiente:

"Se os van a repartir unas hojas de respuesta y un cuadernillo con la portada hacia abajo", se repetirá: "debéis tener cada uno una hoja de respuestas y un cuadernillo de trabajo" a continuación se dirá: "dad primero la vuelta a la hoja de respuestas. Como veis hay unos datos ya rellenados por el profesor. Vosotros ahora debéis rellenar: número, nombre y apellidos, edad, sexo (H si es chico y M si es chica), Centro y curso. ¿Queda todo claro?. Si alguno tiene alguna duda que levante la mano y yo iré personalmente a resolvérsela. Podéis rellenar vuestros datos".... (PAUSA).

Cuando se observe que todos han terminado dirá: "En esta hoja de respuestas debéis contestar a los ejercicios del Cuadernillo marcando la letra que corresponde a la pregunta correcta. Nunca escribáis en el cuadernillo. Dad la vuelta al Cuadernillo de trabajo e id leyendo en voz baja lo que está escrito en la primera hoja mientras yo lo leo en voz alta". El profesor leerá en voz alta, mientras los alumnos lo hacen mentalmente lo escrito en la portadilla del Cuadernillo y en las instrucciones. Se cerciorará de que han comprendido bien los ejemplos. A continuación dirá:

"¿Queda claro cómo debéis trabajar?. Si alguno tiene duda que levante la mano e iré personalmente a resolverla".

Después de cerciorarse de que todos han comprendido el modo de trabajar dirá:

"Como veis contáis con 60 minutos como máximo para hacer la prueba. Pero como vais a trabajar rápidamente y sin equivocarse estoy seguro que la mayoría de vosotros terminará antes. Al acabar los ejercicios me entregáis la Hoja de respuestas y el Cuadernillo, y salís en silencio. ¿Estáis preparado?. Pasad a la página siguiente y COMENZAD!

A partir de este momento pondrá el cronómetro en marcha y empezará el tiempo. Según vayan entregando pondrá a un lado los cuadernillos y anotará el tiempo reducido a minutos en la Hoja de Respuestas. Por ejemplo, si un alumno tarda 20 m 25 seg, dividirá los 25 seg. entre 60 y escribirá 20,42 minutos.

Cuando hayan pasado los 60 minutos, tiempo máximo previsto dirá:

"¡Atención!.¡BASTA!. Dejad los lapiceros sobre la mesa y poned la Hoja de Respuestas y el Cuadernillo con la portada hacia abajo. No os mováis que paso yo a recogerlos".

Se recogen las Hojas de Respuestas y los Cuadernillos y se da por finalizada la Prueba.

Es importante el orden en que se incluyen las preguntas en la prueba, ya que la secuencia de los ítems puede influir en el rendimiento del sujeto. Para ello existen una serie de pautas generales que pueden ser muy útiles en la estructuración de la prueba:

- Al inicio de la prueba se debe incluir una presentación solicitando la cooperación del sujeto especificando los objetivos de la prueba, animando a la contestación de los ítems e incluso informando de la valoración de los mismos y por tanto de la prueba.
- Los primeros ítems deben ser de carácter general, sencillos y potenciadores del interés del sujeto. En pruebas no cognitivas serán también fáciles para lograr que el sujeto se encuentre cómodo al realizar la prueba.
- Los ítems de tipo general deben preceder a las preguntas más específicas y substantivas e importantes de la investigación.

- En prueba no cognitivas las preguntas menos problemáticas deben formularse inicialmente y de forma gradual se introducirán las preguntas más complejas y personales.
- El flujo en el proceso de ítems debe ser lógico de acuerdo con la perspectiva del sujeto y, al mismo tiempo, las preguntas deben agruparse en función de su temática para evitar el desconcierto del sujeto haciendo preguntas relativas a un mismo tema en diferentes fases de la prueba.
- Los datos de identificación del sujeto, utilizados como variables de análisis y a nivel global para comprobar la fiabilidad de la muestra deben formularse al final de la prueba.
- Respecto al aspecto formal de la prueba se debe evitar que un ítem aparezca cortado en la página, si un ítem tiene asociado un gráfico éste debe aparecer con anterioridad al ítem, si se cambia el tipo de ítems en la prueba se debe explicar la forma de cumplimentarlos si resulta problemática la respuesta.

El **registro de los resultados** de la prueba se puede realizar en el propio cuadernillo de la prueba o en hoja de registro preparada "ad hoc". Para utilizarse posteriormente la prueba es aconsejable la segunda opción aunque puede representar cierta dificultad sobre todo si la población objetivo es de corta edad.

Dado el proceso tan largo y laborioso que supone la confección de una prueba es conveniente, si existe, utilizar las pruebas de empresas o entidades dedicadas exclusivamente a estos menesteres (véase TEA, CEPE, Ars Médica, etc.).

5. APLICACIÓN DE LA PRUEBA

La aplicación o administración de la prueba se puede realizar de distintas maneras:

a) **Colectiva o individual**. La aplicación de la prueba se hace al grupo durante la sesión de aplicación o bien se realiza de forma individual al sujeto. Según los requerimientos del test se podrá optar por una de las dos alternativas.

b) Como una **prueba escrita** o como una **prueba en la web**. El creciente desarrollo del ordenador y la telemática facilita la aplicación de los test en un medio informatizado aligerando a su vez la recogida de datos al

grabarse directamente los resultados en una base de datos, una hoja de cálculo o el correo electrónico. La prueba escrita recoge la información directamente en el papel y por tanto, aunque resulta más fácil su cumplimentación, requiere más tiempo para el registro de los datos.

c) **Personal, por correo o por teléfono**. Estas tres formas distintas de aplicar la prueba requieren de la presencia del investigador en el primer caso y de un grupo de personas especialmente preparadas y con un sistema expresamente dedicado a la administración de la prueba por teléfono (sistema CATI). La administración por correo exige mayor cantidad de pruebas a enviar (cuando menos el doble de la cantidad deseada) y especial cuidado en las instrucciones de cumplimentación. De los tres medios, sólo la aplicación por teléfono realizada con el sistema CATI facilita el registro de la información al grabarse los resultados directamente en un fichero.

6. PUNTUACIÓN DE LA PRUEBA

Una vez construida y aplicada la prueba, para hacer operativo el instrumento de medida se necesita:

a) Asignar puntuaciones a los ítems

b) Construir la puntuación de la prueba

6.1. Puntuaciones de los ítems

Según el tipo de ítem así será la manera de asignar la puntuación:

Tabla 4.4. Asignación de numerales a los ítems de la prueba

Tipo de prueba	Tipo de ítem	Valoración (por ejemplo)
Cognitiva	Elección múltiple	2 (respuesta correcta); 1 (respuesta incorrecta)
	Dos alternativas	2 (respuesta correcta); 1 (respuesta incorrecta)
	Emparejamiento	Número de emparejamientos correctos
No cognitiva	Escala de categorías	Asignar un número entero a cada categoría y anotar el número de la categoría escogida. Posición más negativa el 1 e ir subiendo de valor.
	Listado	2 (si marca un término) y 1 (si no lo marca)

En el caso de un ítem de respuesta abierta para cuantificar el resultado necesitará un paso previo de categorización. Este proceso exige la lectura de los ítems y la extracción de categorías y la posterior asignación de numerales. Con la llegada del ordenador y los procesos estadísticos de minería de datos y análisis textuales, la categorización se puede realizar de forma automática en primera instancia aunque finalmente intervenga el investigador. Programas como el Pasw Text Analytics de SPSS o el módulo tm de R realizan la labor detallada e incluso análisis de datos de las categorías construidas.

6.2. Puntuación de la prueba

La puntuación de la prueba habitualmente resulta de sumar las puntuaciones de todos los ítems. Si el investigador desea ponderar algunos ítems la suma será ponderada según estos valores.

En las prueba no cognitivas con diferentes categorías, si los ítems forman una escala sumativa tipo Likert se procederá en primer lugar a comprobar dicho hecho (ver capítulo 7) y luego se sumarán los ítems para obtener el valor de la escala.

En las pruebas cognitivas de elección múltiple se puede dar el caso que los sujetos contesten a los ítems por azar. Para evitar esta actitud se suele corregir la puntuación total mediante la siguiente fórmula:

$$X = A - \frac{E}{K-1}$$

A = ítems acertados en la prueba; E = ítems no acertados; K = número de categorías de los ítems

7. INTERPRETACIÓN DE LAS PUNTUACIONES DE LA PRUEBA

Una vez calculada la puntuación de un sujeto en la prueba se necesita referir dicha puntuación o bien a un grupo normativo o de referencia (norma) o bien según un criterio (prueba criterial). En este capítulo revisaremos el primer grupo y dejaremos al capítulo 8 para el segundo grupo.

La puntuación respecto al grupo normativo puede ser de tres tipos (Navas, 2001, pp. 149-150) puntuaciones **cronológicas**, puntuaciones **centiles** y puntuaciones **típicas**. Las primeras establecen la relación entre la edad cronológica del sujeto y la edad según el grupo normativo. Las segundas indican el porcentaje de sujetos del grupo normativo que están por debajo de la puntuación obtenida por el individuo. Finalmente la puntuación típica traduce la puntuación del sujeto a la puntuación típica que obtendría en el grupo normativo, es decir, hace referencia a la media y la desviación típica del grupo normativo.

Por ejemplo, si en una prueba realizada a estudiantes de 10 años la media obtenida es 30. Si un sujeto realiza la prueba y tiene 30 puntos entonces se le asociara la puntuación de 10 (puntuación cronológica).

El procedimiento para hallar las puntuaciones centiles es el siguiente:

1.º Se aplica el test al grupo normativo y se obtiene la correspondiente distribución de frecuencias.

2.º Se hallan las frecuencias acumuladas.

3.º Se hallan los porcentajes acumulados.

4.º Se hallan los puntos centiles correspondientes a cada puntuación directa, por medio de una interpolación aritmética.

Veamos un ejemplo de la puntuación centíl: sea el resultado de una prueba de matemáticas realizada a un grupo, se quiere saber la puntuación centil que tendrá un sujeto que obtiene en la prueba la puntuación de 18. Según se puede ver en la tabla la puntuación centil será de 46,2, es decir tiene el 46,2% de individuos con puntuación inferior a él.

Tabla 4.5. Puntuaciones del grupo normativo

Puntuación en matemáticas	Frecuencia	Porcentaje	Porcentaje acumulado
10	2	,9	,9
11	3	1,3	2,1
12	6	2,6	4,7
13	10	4,3	9,0
14	12	5,1	14,1
15	15	6,4	20,5
16	18	7,7	28,2
17	20	8,5	36,8
18	22	9,4	46,2
19	24	10,3	56,4
20	21	9,0	65,4
21	18	7,7	73,1
22	16	6,8	79,9
23	14	6,0	85,9
24	12	5,1	91,0
25	8	3,4	94,4
26	6	2,6	97,0
27	4	1,7	98,7
28	2	,9	99,6
29	1	,4	100,0
Total	234	100,0	

Para hallar la puntuación típica del individuo que ha tenido 18 puntos de puntuación directa será:

$$z = \frac{x - \mu}{\sigma}; \quad z = \frac{18 - 18,96}{3,97} = -0,2418;$$

donde 18,96 es la media del grupo normativo y 3,97 su desviación típica.

Y si queremos saber el percentil que le correspondería si la distribución fuera normal, se calculará la probabilidad de z<-0,2418, que si se efectúa dicho cálculo resulta 0,4044, es decir, 40,44%, valor no muy alejado del 46,2 obtenido antes lo que nos da una idea del ajuste de los datos a la distribución normal.

Hay otras escalas transformadas de las puntuaciones típicas, mediante el cambio lineal: y=a+bz, las más utilizadas son:

a) t=50+10z, distribución normal de media 50 y desviación típica 10. Por ejemplo nuestro sujeto tendrá una puntuación t=50+10(-0,2418)=47,58

b) d=50+20z. En nuestro ejemplo será: d=50+20(-0,2418)=45,16

8. CLASIFICACIÓN DE LOS TEST

La cantidad y variedad de test es muy elevada. Atendiendo a diferentes criterios de clasificación, podemos ver en la **tabla 4.6** y **tabla 4.7** una tipología de los mismos.

Tabla 4.6. Clasificación de los test según criterio (Elaborado sobre la base de Martínez Arias, 1995)

CRITERIO	TIPOS
Método	Psicométricos Proyectivos
Finalidad	Investigación Diagnóstico
Área de comportamiento	Rendimiento Inteligencia y aptitudes Personalidad, actitudes e intereses
Modalidad de aplicación	Orales Papel y lápiz Manipulativos o de ejecución Situacionales o simulaciones Computerizados Objetivos
Demandas exigidas al sujeto	Velocidad o rapidez Potencia o dificultad
Modelo estadístico de base	Teoría clásica de los test Teoría de la generalizabilidad Teoría de respuesta al ítem
Modelo conceptual	Teoría referida al criterio Procesamiento de la información Rasgo latente

Tabla 4.7. Clasificación de los test según tipo (Elaborado sobre la base de Bisquerra, 1989)

Tipo	Ejemplos
Rendimiento académico	De cada materia, lectura, escritura, cálculo
Inteligencia	Wais, Wisc, Factor G de Catell, Matrices progresivas de Raven
Aptitudes diferenciales	Factor verbal, factor numérico, razonamiento abstracto, factor espacial, razonamiento mecánico, rapidez y precisión perceptiva, memoria y atención. Ejemplo: DAT
Aptitudes especiales	Habilidad psicomotoras, destreza manual, destreza digital, ritmo, velocidad, artísticas, musicales, etc.
Personalidad	Temperamento, carácter, ajuste, introversión-extroversión, autoconcepto. Ejemplo: MMPI, EPI, 16PF de Catell, etc.
Proyectivos	Psicodiagnóstico de Rorschach, test del árbol de Koch, grafología, etc.
Autoconcepto	Tennessee Self-Concept Scale, etc.
Cuestionario de intereses	Inventario de intereses profesionales
Valores	Cuestionario de valores de Scott, test de reacción valorativa de V. García Hoz, etc.
Sociométricos	Basados en el sociograma introducido por Moreno
Creatividad	Test de pensamiento divergente de Guilford, test de asociaciones remotas de Mednick, etc.

Para estar actualizado en los diferentes test o pruebas es aconsejable consultar a las empresas o entidades dedicadas a estos menesteres. Además en Internet existen multitud de ejemplos de pruebas (p.e http://historia.fcs.ucr.ac.cr/bases/aptitud/aptitud.htm consulta 14-8-2010) y test (p.e. http://www.psicoactiva.com/test/test9.htm consulta 14-8-2010) además de información complementaria sobre la construcción de ítems y formatos de preguntas.

BIBLIOGRAFÍA

AMERICAN PSYCHOLOGICAL ASSOCIATION, AMERICAN EDUCATIONAL RESEARCH ASSOCIATION, AND NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION (1999): Standards for educational psychological test. American Psychological Association. Washington D.C.(http://www.apa.org/science/standards/html)

ANASTASI, A. (1980). *Test psicológicos*. Madrid: Aguilar.

ASSESMENT SYSTEMS CORPORATION (1986). User's Manual for ITEMAN, RASCAL and ASCAL. Minnesota: Assesment Systems Corporation.

BARBERO, M. I. (1993). *Psicometría II. Métodos de elaboración de escalas.* Madrid: UNED.

BISQUERRA, R. (1989). *Métodos de Investigación educativa. Guía práctica*. Barcelona: Ceac.

BLOOM, B. S. (ed.) (1956). *Taxonomy of Educational Objectives, the classification of educational goals – Handbook I: Cognitive Domain*. New York: McKay.

BRIONES, G. (1982). *Métodos y técnicas de investigación para las ciencias sociales*. Mexico: Trillas.

BUGEDA, J. (1975). *Sociología matemática.* Madrid: Instituto de Estudios Políticos.

BUNGE, M. (1976). *La investigación científica*. Barcelona: Ariel.

CALVO, F. (1990). *Estadística aplicada: con el planteamiento y resolución de 450 problemas*. Bilbao: Deusto.

DE KETELE, J. M. y ROEGIERS, X. (1995). *Metodología para la recogida de información*. Madrid: La Muralla.

DEL RINCÓN, D. y otros (1995). *Técnicas de investigación en Ciencias Sociales*. Madrid: Dykinson.

DUVERGER, M. (1974). *Método de las ciencias sociales*. Barcelona: Ariel.

GARCÍA CUETO, E. (1993). *Introducción a la Psicometría*. Madrid: Siglo XXI de España.

GARCÍA RAMOS, J. M. (1989). *Bases pedagógicas de la evaluación. Guía práctica para educadores*. Madrid: Síntesis.

KERLINGER, F. N. (1985). *Investigación del comportamiento. Técnicas y metodología*. México: Interamericana.

López Feal, R. (1986). *Construcción de instrumentos de medida en Ciencias Conductuales y Sociales.* Barcelona: Alamex.

López Pina, J. A. (1995). *Teoría de Respuesta al Ítem: Fundamentos*. Barcelona: PPU

Martínez Arias, M. R. (1995). *Psicometría: Teoría de los Test Psicológicos y Educativos*. Madrid: Síntesis.

Montoya Sáenz, J. M. (1988). *Actitudes sociopolíticas de los estudiantes españoles*. Madrid: UNED.

Morales Vallejo, P. (1988). *Medición de actitudes en Psicología y Educación*. San Sebastián: Tártalo.

Muñiz Fernández, J. (1990). *Teoría de Respuesta a los items*. Madrid: Pirámide.

Muñiz Fernández, J. (2000). *Teoría clásica de los test.* Madrid: Pirámide.

Navas Ara, M. J. (2001). *Métodos, diseños y técnicas de investigación psicológica.* Madrid: UNED.

Nunnally, J. C. (1978). Psychometric Theory. Nueva York: McGraw-Hill (Traducido por Trillas 1987).

Oppenheim, A. N. (1992). Questionnaire design, interviewing and attitude measurement. London; New York: Printed Publishers, cop.

Renom Pinsach, J. (1992). *Diseño de test.* Hospital de Llobregat (Barcelona): IDEA, Investigación y Desarrollo.

Rincón, D. del; Arnal, J.; Latorre, A. y Sans, A. (1995). *Técnicas de investigación en Ciencias Sociales.* Madrid: Dykinson.

Santisteban, C. (1990). *Psicometría. Teoría y práctica en la construcción de test*. Ediciones Norma. Madrid.

Sierra Bravo, R. (1985). *Técnicas de investigación social. Teoría y ejercicios*. Madrid: Paraninfo.

Silva, F. (1989). *Evaluación Conductual y Criterios Psicométricos*. Madrid: Pirámide.

Summers, G. F. (1976*). Medición de actitudes*. México: Trillas.

Visauta, S. (1989). *Técnicas de investigación social*. Barcelona: PPU.

Walker, R. (1989). *Métodos de investigación para el profesorado.* Madrid: Morata.

Yela, M. (1984). *Introducción a la teoría de los test*. Madrid: Facultad de Psicología, Universidad Complutense.

Capítulo 5
TÉCNICAS DE ENCUESTAS

1. Introducción
2. Tipos de encuestas
   2.1. La encuesta personal
   2.2. La encuesta telefónica
   2.3. La encuesta postal
   2.4. Las encuestas por ordenador y on-line
3. Criterios en la selección del tipo de encuesta
4. Proceso de una encuesta
   4.1. Diseño de cuestionarios
   4.2. Diseños muestrales
   4.3. El trabajo de campo
   4.4. El análisis estadístico
   4.5. El informe de resultados
   4.6. Ayudas informáticas para el tratamiento de encuestas
   4.7. Conclusiones
5. Un ejemplo de encuesta

Objetivos

- Valorar la utilidad de la encuestas en las investigaciones sociales
- Conocer las distintas variantes de encuestas y sus limitaciones
- Saber elegir el modelo de encuesta más apropiado a un contexto concreto
- Diseñar un cuestionario
- Realizar los controles técnicos apropiados para controlar la calidad de la encuesta
- Saber implementar una encuesta
- Recoger y analizar los datos emanados de una encuesta

1. INTRODUCCIÓN

Algunos definen la investigación mediante encuestas por: técnica cuantitativa que consiste en una investigación realizada sobre una muestra de sujetos, representativa de un colectivo más amplio que se lleva a cabo en el contexto de la vida cotidiana, utilizando procedimientos estandarizados de interrogación con el fin de conseguir mediciones cuantitativas sobre una gran cantidad de características objetivas y subjetivas de la población. El **Cuestionario** es el instrumento de recogida de datos de la encuesta.

Ventajas:

— Técnica más utilizada y que permite obtener información de casi cualquier tipo de población.

— Permite obtener información sobre hechos pasados de los encuestados.

— Gran capacidad para estandarizar datos, lo que permite su tratamiento informático y el análisis estadístico.

— Relativamente barata para la información que se obtiene con ella.

Inconvenientes:

— No permite analizar con profundidad temas complejos (recurrir a grupos de discusión).

2. TIPO DE ENCUESTAS

Existe diferentes criterios de clasificación, no obstante se presenta el de mayor consenso:

- Por la forma de realizarse:

  - Cara a cara
  - Postal
  - Telefónica
  - Por ordenador
  - On-line
- Por su contenido:
  - Referidas a hechos
  - Referidas a opiniones
  - Referidas a actitudes, motivaciones o sentimientos
- Por la naturaleza de su aplicación:
  - Puntuales o transversales
  - Panelísticas o longitudinales
- Por sus fines científicos:
  - Exploratorias
  - Descriptivas
  - Explicativas
  - Predictivas
  - Evaluativa
- Por su finalidad:
  - Político-sociales
  - Comerciales
  - Con fines específicos

Por la forma de realizarse la entrevista, personal o encuesta cara a cara, hacen uso de encuestadores. La encuesta por correo consiste en el envío por correo de un cuestionario, es más barata, pero tienen el inconveniente de un índice de respuesta no elevado, por lo que hay que hacer sucesivas oleadas, lo que puede hacer que la muestra no sea representativa. En los cuestionarios telefónicos no controlamos a la persona que responde pero son econó-

micas de realizar. Los cuestionarios autoadministrados o por ordenador, se suelen realizan a una población cautiva.

2.1. La encuesta personal

La encuesta personal es quizás el método que goza de mayor popularidad y el que se ha utilizado con mayor profusión en la captación de información primaria debido, principalmente, a las ventajas que presenta frente a los otros tipos de encuesta. Consiste en una entrevista personal que se establece entre dos personas, a iniciativa del entrevistador, para obtener información sobre unos objetivos determinados.

Las principales ventajas de las encuestas realizadas personalmente son las siguientes:

1. Entre los distintos métodos de encuesta, las personales son las que proporcionan un mayor índice de respuesta, ya que cuando se contacta con las personas a encuestar es poco probable que declinen responder el cuestionario o no lo concluyan una vez comenzado.
2. Son fiables puesto que se conoce con certeza quién contesta y se evita la influencia de terceras personas.
3. Se obtienen respuestas menos evasivas e inconcretas, ya que el entrevistador puede aclarar cualquier tipo de dudas que se puedan suscitar en el cuestionario y se reducen de forma considerable las típicas respuestas de no sabe no contesta.
4. Permiten utilizar materiales auxiliares para profundizar en determinados temas, tales como fotografía, láminas, productos, etc.
5. También pueden obtenerse datos secundarios del entrevistado como presencia, ambiente familiar, sexo, etc.

En contrapartida, las encuestas personales presentan como principales inconvenientes:

1. Un coste elevado, ya que al tiempo necesario para realizar la entrevista hay que añadirle el tiempo que el entrevistador necesita para desplazarse hasta el lugar de residencia de la persona a entrevistar que, junto con el coste de los viajes, representa un coste importante por entrevista realizada.

2. Pueden originarse sesgos por influencias del entrevistador, es decir, la actuación del entrevistador en la encuesta, su presencia física, sus explicaciones, su entonación, etc. pueden representar una fuente de distorsión en la información obtenida. Para evitar estos posibles sesgos, es importante que en la selección los entrevistadores se tengan en cuenta ciertas cualidades personales, honestidad, objetividad, neutralidad y facilidad para los contactos humanos, siendo también importante una buena formación de carácter general sobre la técnica de la entrevista, formas de presentación, aclaración de respuestas, etc, además de las instrucciones concretas de cada investigación en particular que aunque traten sobre el mismo tema emplean cuestionarios y procedimientos de búsqueda de las unidades muestrales diferentes.
3. Teniendo en cuenta que la base de la investigación descansa en la veracidad de la información recogida, es necesario controlar que los datos presentados por los entrevistadores son ciertos y que proceden realmente de la muestra seleccionada. Para verificar esto, en las encuestas personales es necesario un exhaustivo control de los mismos, a través de una revisión de una pequeña muestra de las encuestas realizadas por cada encuestador.

En lo que a la realización de las encuestas personales se refiere pueden llevarse a cabo bien en el hogar de la persona encuestada o "in situ". Las encuestas efectuadas en el hogar se utilizan en investigaciones de tipo general, garantizándose la aleatoriedad de los hogares seleccionados a través de diferentes métodos de muestreo. Las encuestas "in situ" se realizan en diferentes lugares relacionados con el objetivo de la investigación (establecimientos estaciones de autobuses, etc.) con la finalidad de efectuar la encuesta en una situación más real.

2.2. La encuesta telefónica

La encuesta telefónica es un método cuya utilización va en aumento en los últimos años a medida que se incrementa el número de hogares con teléfono y mejora, por tanto, su representatividad. Inicialmente se utilizó para realizar test de audiencia de programas y anuncios emitidos por radio y televisión, pero posteriormente se ha generalizado su uso en la captación de información. En investigación de mercados la encuesta telefónica puede

utilizarse en tres sentidos: 1) como único medio de realizar la investigación, 2) como una técnica combinada con otras, es decir, se puede diseñar una encuesta para realizarse personalmente, telefónicamente y por correo en función de las características y condicionantes de la investigación y de la población de la cual se pretende obtener la información y 3) como un medio auxiliar en la utilización de otras técnicas como, por ejemplo, para establecer citas en la realización de encuestas personales, informar o reclamar el envío por correo de cuestionarios, completar cuestionarios, realizados a través de encuestas postales, como medio de control de los encuestadores en entrevistas personales o telefónicas, etc.

Entre las principales ventajas que presenta la encuesta telefónica podemos destacar las siguientes:

1. Es una técnica que permite obtener información con gran rapidez, ya que en pocas horas un encuestador puede realizar múltiples encuestas.
2. En comparación con la encuesta personal su coste puede resultar más reducido, aunque esto depende de diversos factores tales como la duración del cuestionario, la distancia, las tarifas telefónicas, etc.
3. Permite acceder más fácilmente a personas ocupadas, sobre todo para encuestas de corta duración, a las cuales suele ser difícil llegar a través de encuestas personales.
4. En relación al índice de respuesta, el número de personas que aceptan contestar por teléfono una encuesta suele ser bastante elevado, aunque en los últimos años la rápida expansión de Telemarketing "dirigido" está provocando una disminución de la disposición del público a cooperar en las encuestas telefónicas como consecuencia del aumento de llamadas telefónicas, muchas veces indeseadas, y la confusión que puede producirse en los individuos entre el telemarketing y los estudios de investigación de mercados. Por ello, es fundamental dar una explicación clara y convincente de que se trata de un estudio de investigación comercial, sin ninguna intención de vender nada, informarle del objetivo de la investigación, quienes realizan y de cómo esa persona ha sido seleccionada.

Los principales inconvenientes o limitaciones de la encuesta telefónica son:

1. Los posibles problemas de falta de representatividad muestral como consecuencia de que sólo pueden ser encuestadas las personas que disponen de teléfono, si bien en el caso de que el universo lo constituyan las empresas no existe este problema pues prácticamente el cien por cien de ellas poseen teléfono. Con respecto a la selección de los números telefónicos es conveniente generarlos aleatoriamente a través de un directorio telefónico pues hay números que no están registrados y no incluye los nuevos abonados.

2. Otro de los aspectos que supone una limitación de este método es la brevedad de la entrevista que se recomienda que no exceda de 15 minutos. Ello implica que el cuestionario también sea breve y, por tanto, no puede recabarse demasiada información.

3. El no poder utilizar material auxiliar o de exhibición, como tarjetas, fotografías, bocetos, etc. limita también bastante el tipo de información que se puede obtener a través de las encuestas telefónicas.

En relación al diseño del cuestionario de las encuestas telefónicas hemos de tener presentes, además de las reglas generales que más adelante se comentarán, los siguientes aspectos:

a) La brevedad de la entrevista exige que el cuestionario sea reducido y no contenga demasiadas preguntas.

b) Las preguntas deben ser preferentemente cerradas o semi-abiertas y con un abanico de respuestas no muy amplio.

c) Las preguntas de respuesta múltiple se deben eliminar o desglosar en sucesivas preguntas.

2.3. La encuesta postal

La encuesta postal consiste en el envío por correo de un cuestionario a las personas que constituyen la muestra con la esperanza de que por la misma vía lo devuelvan cumplimentado. Como técnica de entrevista en algunas ocasiones puede tener una utilización esencial cuando es el único medio de establecer contacto con los encuestados y en otras situaciones puede utilizarse como técnica opcional en función de las características de la investigación. Entre las principales ventajas de la encuesta postal hemos de destacar:

1. Su reducido coste frente a los otros dos método de encuesta, pues es un sistema económico de recogida de información especialmente cuando el ámbito geográfico de la investigación es a nivel nacional o internacional.
2. Es un sistema más flexible para el entrevistado que puede contestar el cuestionario en el momento más conveniente y empleando el tiempo que desee, lo cual puede llevar a respuestas más meditadas.
3. Se evitan los posibles sesgos que puedan originarse por la influencia del encuestador en el proceso de realización de la encuesta.

En lo que se refiere a las desventajas e inconvenientes de la encuesta postal destacamos como más importantes:

1. El bajo número de respuestas que se reciben en relación con los envíos realizados, que se sitúa en torno al 10% o 20%. Para tratar este problema del bajo índice de respuesta es necesario establecer una serie de mecanismos motivadores, como son, en primer lugar que junto con el cuestionario se envíe una carta de representación en la que se ponga de manifiesto la necesidad de la colaboración del encuestado y la importancia que tiene su respuesta para los fines de la investigación. En segundo lugar, otro mecanismo que incrementa el índice de respuesta es ofrecer un incentivo material, además del incentivo moral que se pretende dar con la carta. En este sentido, se han utilizado diversas formas como, por ejemplo, ofrecer una participación en sorteos a las personas que respondan, regalos, vales de descuento sobre determinados productos, facilitarlos resultados de la investigación, etc.
2. La identidad del encuestado es una variable incontrolada, ya que no podemos asegurar que realmente conteste el cuestionario el destinatario del mismo y, al mismo tiempo, también pueden existir influencias de otras personas en las respuestas del encuestado y ello, lógicamente, introduce distorsiones en la muestra.
3. Las encuestas postales están limitadas a situaciones en las que se dispone de un directorio del universo a investigar, lo cual no siempre es frecuente en las investigaciones.
4. Otro de los problemas que motiva la ausencia de un entrevistador hace referencia a la poca cantidad de información que se puede obtener a través del cuestionario, que necesariamente ha de ser breve,

salvo que los encuestados estén emocionalmente involucrados en el tema a investigar. Si bien no existe una regla general sobre la extensión del cuestionario, normalmente la brevedad del mismo incrementa el número de respuestas.

5. Finalmente, hemos de señalar los problemas asociados a la representatividad de la muestra, ya que la falta de respuesta puede dar lugar a que la estructura de la muestra real que se obtiene de los cuestionarios contestados no se ajuste a estructura de la muestra definida, siendo necesario o bien enviar nuevos cuestionarios a los estratos de la muestra de los que nos e han recibido suficientes respuestas, o completar las respuestas necesarias mediante entrevistas telefónicas o personales o bien desechar un determinado número de encuestas de aquellos estratos de la muestra cuyas respuestas fueron más numerosas, incrementándose en este último caso el error muestral.

Tal y como señalamos anteriormente, el envío de un cuestionario postal debe ir acompañado de una carta de presentación con la que se pretende estimular a los encuestados a cumplimentar y devolver la encuesta y, al mismo tiempo, sirve como medio de comunicación entre el Investigador y el encuestado para dar a conocer los motivos de la investigación, quién la realiza, sus objetivos, etc.

En la presentación y redacción de esta carta deben tenerse presentes los principios básicos de la publicidad directa, es decir:

a) debe estar dirigida normalmente al destinatario

b) redactada desde el punto de vista del destinatario y no del remitente

c) el principio y el final de la carta son las dos partes más importantes, por lo que se deben incluir aspectos atractivos, evitando introducciones innecesarias y largas despedidas finales

d) debe ir firmada en original

e) no debe ser excesivamente larga, normalmente se recomienda que su extensión no supere un folio

En relación al contenido de la carta debe hacerse referencia expresa a los siguientes aspectos:

a) quién realiza la investigación,

b) objetivos y repercusiones,

c) necesidades de colaboración del entrevistado,

d) anonimato del encuestado y tratamiento global de los datos facilitados.

El cuestionario, tanto en su forma como en su contenido, debe ser fácil de contestar, las preguntas deben estar formuladas de una forma clara y concisa, y deben ser cerradas o semi-abiertas, procurando que la respuesta no sea ambigua. Como parte integrante del cuestionario o en documento separado se deben incluir unas instrucciones breves, claras y precisas, para facilitar su cumplimentación. Es conveniente, además, que los cuestionarios estén numerados, con el objeto de que si después de un primer envío no se obtiene respuesta, se puede repetir, enviando un segundo cuestionario. Como es obvio, el envío debe contener también un sobre franqueado con la dirección del investigador para la devolución del cuestionario cumplimentado.

2.4. Las encuestas por ordenador y on-line

Las encuestas por ordenador tienen una variante cuyo desarrollo está en continuo crecimiento, son las encuestas on-line o por Internet, por tanto haremos especial incidencia en éstas últimas al tener mayor aplicación.

Las encuestas on-line tienen tres fases perfectamente definidas: crear el cuestionario, publicar en la web el mismo y finalmente administrar el citado cuestionario.

La creación del cuestionario

El cuestionario en la web tiene que tener unas características diferencias:

a) Debe ser lo más amigable posible, con un diseño atractivo, y despertar un interés para retener al encuestado hasta el final del cuestionario

b) En su diseño se debe tener en cuenta la compatibilidad de la encuesta con los principales navegadores de Internet (Internet Explorer, Firefox, Chrome, etc.). En la medida de lo posible, se debe evitar el uso de documentos adjuntos, así como la descarga de software adicional para la realización de la encuesta

c) El software utilizado en la creación de la encuesta debe permitir preguntas filtro, de salto y aleatorización del orden de los ítems que componen el cuestionario, etc.

d) Las primeras pantallas deben ser de invitación a contestar y con preguntas fáciles o filtro para que si el encuestado desea abandonar la encuesta no necesite navegar más tiempo. Se debe colocar un puntero para facilitar al encuestado la rectificación de preguntas o el avance en preguntas que no desee contestar

e) Es importante que todas las categorías de respuesta sean visibles, evitando tener que utilizar el ratón para desplazamiento horizontal o vertical

f) Conviene poner una "barra de progreso" que informe al encuestado sobre el punto del cuestionario en que se encuentra, disminuyendo la tasa de abandono, en la medida que el usuario se siente informado sobre la duración de la misma

g) Resulta interesante que el encuestado pueda avanzar en el cuestionario pese a no haber contestado a alguna pregunta. Una pregunta sólo debe ser de respuesta obligada si resulta imprescindible en el desarrollo de la encuesta

h) El empleo racional de imágenes, gráficos o vídeos son recursos motivadores que pueden contribuir a aumentar la calidad de las encuestas y facilitar la cumplimentación de las mismas

La publicación del cuestionario

La publicación del cuestionario representa el alojamiento del mismo en un servidor de Internet donde el encuestado tiene acceso. Esta publicación se puede materializar en una página web, en un bloc, e incluso el sistema puede mandar la encuesta para cumplimentarse en una PDA.

Para acceder a la encuesta se puede realizar mediante un link de acceso exclusivo para cada usuario, de modo que se garantice la unicidad de las respuestas, esto es, que cada encuestado pueda contestar en una sola ocasión; o mediante un link común a todos los encuestados pero con una palabra de paso única para cada usuario.

No se debe fijar una fecha límite de recogida de datos aún cuando se tenga previsto la recogida de los mismos en una fecha límite. Como mínimo se debe fijar un plazo de tres días, pues se ha comprobado que más del 75% contestan en los dos primeros días de recibir el mensaje de invitación.

Debe existir una invitación a cumplimentar la encuesta mediante envío de correo electrónico al encuestado.

La administración del cuestionario

Básicamente la administración del cuestionario tiene encomendadas las tareas de gestión del cuestionario, gestión de los usuarios y la recogida y en algunos casos tratamiento estadístico de los datos.

Como indica Groves (1989), en cualquier encuesta existen cuatro fuentes potenciales de error: **el error de muestreo, el error de medida, el error de no-respuesta y el error de no-cobertura**.

Respecto a las encuestas online, el error de medida hace referencia al resultado de obtener respuestas inadecuadas por culpa de un cuestionario mal administrado. Empero la metodología online supone un avance en la reducción de este error, en tanto que desaparece el sesgo del entrevistador.

El error de cobertura hace referencia a las diferencias socio-demográficas existentes entre los internautas y la población general, y la posibilidad de que una muestra realizada a un colectivo de Internet pueda ser representativa del resto de la población. El error de cobertura se minimizará a medida que aumente la penetración de Internet.

Por otra parte, el error de muestreo es el resultante de realizar una encuesta a un subgrupo de la población en lugar de realizarlo al total de individuos. Actualmente, el muestreo online se realiza mediante las cuotas que fija el investigador.

Finalmente, el error de no respuesta es el derivado del hecho de que no todos los individuos seleccionados de una muestra acaben participando en el estudio. Una administración eficaz del cuestionario así como la captación incentivada de la muestra, son los dos elementos fundamentales para conseguir una elevada tasa de respuesta.

Existe numerosos programas comerciales para realizar las tres fases de creación, publicación y administración de la encuesta, la mayoría son comerciales, no obstante destacamos dos de software libre en español: LimeSurvey y Opina: gestor de encuestas. Ambos programas necesitan ser alojados en un servidor para realizar todas sus funcionalidades. Además está el programa FormLogix que no precisa la descarga y el alojamiento de ningún programa al estar ya alojado en el servidor de la empresa distribuidora.

Google también participa con sus servidores en la creación, publicación y administración de una encuesta, aunque en esta última fase sólo se pueden realizar estadísticas con el fichero Excel donde se recogen las contestaciones de los encuestados. La utilidad de Google para estos menesteres es Google docs y se necesita tener una cuenta en gmail para acceder a la creación de un cuestionario.

3. CRITERIOS EN LA SELECCIÓN DEL TIPO DE ENCUESTA

Una vez expuestas las anteriores consideraciones sobre los diferentes métodos de captar información a través de las encuestas ad – hoc, el problema de la elección idónea para la investigación depende, lógicamente, de los objetivos de la misma y de su adecuación en función de las ventajas y limitaciones que hemos señalado para cada uno de los métodos. Así, por ejemplo, estableciendo como criterios la versatilidad, el tiempo de obtener la información, el coste, la precisión y la conveniencia para el encuestado podemos jerarquizar la elección de estos tres métodos.

En relación a la versatilidad, factores como el mayor o menor número de preguntas que pueden incluirse en el cuestionario, la cantidad de datos que pueden obtenerse, el tipo de preguntas y la posibilidad de utilizar ayudas visuales como tarjetas, fotografías, etc. determinan que la encuesta personal sea mucho más flexible y permita obtener una mayor variedad de información que las encuestas telefónicas o postales.

Cuando el tiempo es un factor clave en la investigación, la encuesta telefónica, lógicamente, es la forma más rápida de obtener la información y si bien en la encuesta personal se puede incrementar el número de encuestadores para suplir las limitaciones de tiempo, en la práctica los problemas asociados con el adiestramiento, la coordinación y el control de un grupo excesivamente numeroso hacen que no sea ni factible ni económicamente

rentable aumentar el número de encuestadores. En la encuesta postal resulta imposible acortar el tiempo que transcurre desde que se envían los cuestionarios hasta que se devuelven cumplimentados.

Con respecto al coste y en términos generales, la encuesta postal es la más económica, y la personal la más costosa. No obstante, puede haber excepciones en función del tamaño del cuestionario, lugar de residencia de los individuos que constituyen la muestra, tarifas telefónicas, etc.

En lo que a la precisión del método utilizado se refiere, no cabe duda que existen una gran cantidad de factores que determinan la mayor o menor precisión de las preguntas, la mayor o menor disposición de los encuestados a contestar, etc. Ahora bien, existen una serie de aspectos que afectan a la exactitud de los datos y difieren en cada uno de los métodos utilizados como son el grado de control sobre la muestra, que se refiere a la capacidad de identificar y obtener los datos de una muestra que sea representativa de los estratos de la población o universo y la capacidad de asegurar la cooperación de los encuestados contactados, el control en la supervisión, es decir, la posibilidad de minimizar preguntas en la encuesta que no se comprendan. En este sentido, la encuesta personal es la que permite obtener un mayor control sobre la muestra y superar los problemas asociados a las dificultades que pueda tener el encuestado a la hora de responder ciertas preguntas, mientras que la encuesta postal, lógicamente, es la que da lugar a menos errores derivados del propio entrevistador.

Finalmente, la encuesta postal, seguida de la telefónica es la más flexible desde la perspectiva del encuestado, pues le permite cumplimentar el cuestionario en el momento más conveniente y empleando el tiempo que desee.

Obviamente, el problema de la elección del método no solo depende de estos cinco criterios y además es muy difícil que en el planteamiento de una investigación un método sea el mejor para todos los criterios. Por tanto, el más adecuado será el que mejor se adecue a las necesidades de información de la investigación teniendo en cuenta las limitaciones de tiempo y de coste. Por otra parte, hemos de destacar que estos tres métodos de captar información a través de encuestas ad-hoc no son mutuamente excluyentes, sino que pueden combinarse y obtenerse así mejores resultados en el diseño de la investigación en función de las ventajas de cada uno de ellos.

4. PROCESO DE UNA ENCUESTA

La investigación mediante encuesta comprende una serie de fases:

a) Definición del estudio: fines, objetivos y concreción de las hipótesis de investigación.

b) Selección y definición de las variables: en correspondencia con los objetivos planteados. Denominación, explicitación de su correspondencia con los objetivos, tipo de variables, valores que toman, etc.

c) Diseño del cuestionario: elección del tipo de cuestionario, diseño de preguntas, contenido de las mismas, organización y redacción del cuestionario.

d) Validación y prueba del cuestionario: trabajo de campo, prueba piloto (pretest), estudio de las características técnicas y cuestionario final.

e) Selección de la muestra: según la finalidad del estudio, limitaciones temporales, margen de error permitido, relación coste eficacia.

f) Aplicación del cuestionario: recogida de datos.

g) Tratamiento de datos: análisis estadístico (descriptivo, inferencial, relacional, multivariante).

h) Elaboración del informe.

Los dos primeros aspectos son comunes a cualquier proceso de investigación y por tanto se suponen conocidos por el lector, en este capítulo nos centramos en los aspectos concretos relacionados con la técnica de investigación mediante encuestas.

4.1. Diseño de cuestionarios

Un cuestionario es algo más que una simple lista de preguntas. Su diseño es más complicado de lo que puede parecer a primera vista. Debe estar cuidadosamente elaborado en forma y contenido, y han de valorarse toda una serie de elementos que lo doten de rigurosidad y sistematicidad.

La fase de diseño del cuestionario, y concretamente en los aspectos referidos al diseño de preguntas (qué y cómo preguntar, determinación de las

escalas de medida de actitudes y respuestas, formatos de presentación, etc.), constituyen el elemento básico del cuestionario y de su diseño adecuado dependerá el éxito de una información representativa de la investigación. Las preguntas en los cuestionarios nos suministrarán una amplia información alusiva a las opiniones, intenciones, juicios, motivos, hábitos y expectativas de los encuestados.

No es posible crear una receta para diseñar o elaborar un cuestionario. Sin embargo, sí es posible dar algunos consejos prácticos para evitar grandes errores. En el momento del diseño del cuestionario hemos de plantearnos una serie de cuestiones: definición del tipo de cuestionario que vamos a utilizar; valoración del tipo de preguntas que debemos incluir; adopción de un criterio adecuado de codificación de las preguntas para el posterior tratamiento de éstas; elección de un tamaño representativo de la muestra sobre la que se pasará el cuestionario; organización y presentación del cuestionario, etc.

4.1.1. Reglas para la formulación de pregunta

A la hora de desarrollar las preguntas del cuestionario que estemos diseñando, se plantean dos problemas: cómo preguntar y qué tipo de preguntas se han de elegir para el estudio. Como respuesta al primer interrogante podemos afirmar, de modo general, que la calidad de una investigación se encuentra condicionada en gran medida por lo acertado que sea el diseño de las preguntas del cuestionario. Si éstas están mal definidas, son ambiguas, o no responden a la naturaleza del análisis, los resultados obtenidos serán estériles. Como se puede observar, preparar un cuestionario resulta más difícil de lo que parece a simple vista. En líneas generales, podemos formular unas reglas básicas para su confección:

- Si se quieren respuestas concretas las preguntas deben ser concretas.
- No deben hacerse preguntas que contengan potencialmente las respuestas.
- Las preguntas que admiten varias respuestas deben ser definidas cuidadosamente.
- El número de categorías de las respuestas es también muy importante. Raramente son necesarias más de 5 ó 7 categorías en preguntas de

actitud o de opinión. Cuando sean necesarias más categorías, es preferible desglosar la pregunta en varias con menos categorías cada una.

- Las preguntas deben ser neutrales. No deben incorporar juicios, opiniones o valoraciones.
- El encuestado no debe esforzarse en recordar ante lo que conviene facilitar su memoria.
- El orden del cuestionario deberá ser lógico. Las preguntas deben hacerse por temas afines y en orden de dificultad creciente. Las preguntas más sencillas han de ir al principio del cuestionario.
- El orden de las preguntas no debe afectar a las respuestas.
- Los identificadores del encuestado deben ir al final del cuestionario.
- No deben ser excesivamente largo, porque en cuestionarios largos (+100 preguntas) disminuye el % de respuestas.
- Tiene que ser sencillas y redactadas de tal forma que puedan comprenderse con facilidad (no utilizar términos técnicos).
- No deben incorporar términos morales (juicios de valor).
- Nunca sugerir la respuesta, incitando a contestar más en un sentido que en otra.
- Todas las que estén dentro de un mismo tema deben ir juntas en el cuestionario en forma de batería.
- No juntar preguntas cuya contestación a una de ellas influya sobre la contestación del otro, denominado efecto “halo”.
- Debe utilizarse un lenguaje accesible, que se entienda, es decir, el vocabulario ha de ser sencillo, directo, y familiar. No hay que olvidar que salvo estudios específicos, las preguntas van dirigidas a personas de las más diversas clases y formación y consecuentemente hay que tenerlo en cuenta excluyendo la utilización de palabras o conceptos difíciles.
- No se deben utilizar palabras o frases ambiguas que puedan tener distintas interpretaciones. Así por ejemplo, el término “frecuentemente” en la utilización de un método de trabajo puede significar todos los días, dos o tres veces en semana, etc.

- Las preguntas deben ser neutras o imparciales, evitando que de alguna forma pueda estar implícita una determinada respuesta. Por ejemplo, a la pregunta ¿está usted a favor de que el Ayuntamiento de Madrid cree más zonas verdes en la ciudad? la totalidad de los entrevistados contestarán positivamente y estas respuestas no coincidirían si la pregunta se formulase de la siguiente forma: ¿está usted a favor de la creación de más zonas verdes en la ciudad aunque ello suponga un incremento de los impuestos por parte del Ayuntamiento?
- Se deben evitar la utilización de palabras cargadas de ciertas connotaciones, puesto que en muchas ocasiones una pregunta formulada de una determinada forma puede sugerir la respuesta en una determinada dirección.
- No se deben incluir preguntas que sean difíciles de contestar o requieren cálculos, o esfuerzos memorísticos por parte del encuestado y probablemente no los recuerde todos, es preferible enumerarle los productos o servicios y preguntarle si los utiliza o no.
- Tampoco deben formularse preguntas de doble efecto, es decir preguntas cuya redacción pueda implicar una doble respuesta.

Hay preguntas que se refieren a temas delicados o sensibles que pueden poner al entrevistado en una situación embarazosa o preguntas que tienen un elemento de prestigio o de adhesión a normas socialmente aceptadas, que generalmente dan lugar a respuestas con un alto índice de parcialidad. Para evitar los posibles riesgos que puedan producirse en este tipo de preguntas o renuncias del encuestado a contestarlas se han desarrollado deferentes procedimientos para obtener respuestas más exactas:

- A través de enunciados contrarrestados que consisten en iniciar la pregunta con una afirmación que sugiere que el comportamiento en cuestión es común entre la población, y posteriormente, formularle la pregunta.
- Por medio de enunciados indirectos, por ejemplo formulando la pregunta sensible pero referida a otras personas, ya que supone que el comportamiento o la actitud del encuestado se verá reflejada en la respuesta.
- Realizando la pregunta con tarjeta de forma que el encuestado responda con el número que hace referencia a la respuesta en la tarjeta.

- A través de baterías de preguntas que nos permitan acercarnos al tema de una forma indirecta.
- Las preguntas no deben ser excesivamente largas ya que los cuestionarios deben ser fluidos y con preguntas cortas en donde se pueda imponer un ritmo de preguntas y respuestas que no aburran al encuestado ni al encuestador. Cuando este ritmo se altera con una pregunta cuyo texto es excesivamente largo o repetitivo, la conversación pierde dinamismo e incide negativamente en el desarrollo de la entrevista.
- La redacción de las preguntas debe invitar a colaborar, debe formularse de forma amable, natural, y con un lenguaje normal. El entrevistado no debe tener la sensación de estar sometido a un examen, por lo que el cuestionario tiene que diseñarse de forma que entre el encuestado y el encuestador se produzca una situación lo más parecida a una conversación normal.

4.1.2. Tipos de preguntas

El elemento básico del cuestionario, como se deduce claramente de su nombre, son la preguntas. Es por ello que la bondad de un cuestionario depende de la clase de preguntas empleadas en él y de su adecuada formulación. Las preguntas del cuestionario se subdividen en respuestas, que son los elementos de variación o categorías de la variable a que se refiere la pregunta. Existe una amplia tipología de preguntas y diversas formas de clasificación; por ello y sin ánimo de ser exhaustivo nos centraremos en los diferentes tipos de preguntas:

a) *Según la contestación que admitan*:

- *P. abiertas*: Son aquellas en las que no se establece ningún tipo de respuesta, dejando ésta al libre arbitrio del encuestado. Es decir, la respuesta del encuestado no está previamente definida y el encuestador se limita a registrar al pie de la letra la contestación obtenida.
- *P. Cerradas*. Son las que el encuestado se limita a elegir una o varias de las respuestas definidas previamente en el cuestionario; las respuestas se conocen a priori y están totalmente pre-codificadas.

Ambos tipos de preguntas tienen sus ventajas y limitaciones. Con las preguntas abiertas siempre se pueden descubrir nuevas respuestas y opiniones que no se habían tenido en cuenta y se consigue evitar que las respuestas obtenidas puedan estar sesgadas por la línea de hipótesis del investigador. Sin embargo, para poder analizar cuantitativamente este tipo de preguntas es necesario agrupar las respuestas y codificarlas con posterioridad, lo cual entraña cierta dificultad y lleva tiempo. Por otra parte, también presentan el inconveniente de que el entrevistado pueda responder en una línea que no tenga interés para la investigación o incluso que no se adecue a la pregunta. De todos modos, son especialmente adecuadas en investigaciones exploratorias o cuando no se tiene mucho conocimiento sobre las respuestas posibles.

Las preguntas cerradas, por el contrario son más fáciles de contestar dado que requieren un menor esfuerzo por parte del encuestado y lógicamente, no es necesario ni agruparlas ni codificarlas con posterioridad. Por ello, suelen ser las preguntas más utilizadas en los cuestionarios.

Una alternativa intermedia entre las preguntas cerradas y abiertas y de uso frecuente en los cuestionarios es la utilización de preguntas semi-abiertas, es decir preguntas cerradas con un ítem abierto para reservar la posibilidad de incorporar otras respuestas diferentes de las previamente seleccionadas.

En las preguntas cerradas, a su vez, podemos distinguir entre:

- *P. dicotómicas*: Son las que tienen dos únicas respuestas. "Si o No" y a veces Ns/Nc). Se deben utilizar sólo para temas muy bien definidos que admiten estas dos alternativas como respuesta.
- *P. de respuestas categorizadas*: Son aquellas en las que el encuestado debe elegir entre un determinado número de respuestas posibles. En este tipo de preguntas, cuando las posibles opciones de respuesta son numerosas es conveniente hacerlas con tarjeta en vez de leer las respuestas al objeto de impedir que unas respuestas tengan más probabilidad de ser elegidas que otras, ya que cuando el número de respuestas es relativamente grandes, las últimas tienen más probabilidad de ser recordadas que las primeras. Este tipo de preguntas con tarjetas también es conveniente utilizarlo en preguntas con respuestas difíciles o que puedan producir cierto rechazo el contestarlas. Este tipo de preguntas a su vez se divide en:

— *De respuesta espontánea*: el encuestador no debe leerle la respuesta al encuestado.

— *De respuesta sugerida*: el entrevistador lee las preguntas al encuestado.

— *De valoración*: el entrevistador lee una escala de intensidad creciente o decreciente de categorías de respuesta. Estas preguntas de valoración presentan unas variantes:

– *P. de escala subjetiva*: Son aquellas preguntas en que las respuestas se gradúan en intensidad creciente o decreciente sobre el punto de información deseado. Es decir, el encuestado se posiciona subjetivamente respecto a las diferentes categorías de respuesta. (¿Qué le parece...? Bien Mal Peor)

– *P. de escala subjetiva numérica*: Son similares a las anteriores pero con posiciones numéricas.

– *P. Cuadro*: Se utilizan principalmente para obtener más de una información que se recoge normalmente en cuadros de doble entrada.

b) *Según su función en el cuestionario*:

- *Filtro*: se utilizan mucho en los cuestionarios para eliminar aquellas personas que no les afecten determinadas preguntas, es decir que marcan la realización o no de preguntas posteriores. En definitiva, este tipo de preguntas constituyen una bifurcación en el cuestionario.
- *Batería*: todas las preguntas tratan sobre un mismo tema. Siempre deben ir juntas en el cuestionario en forma de batería, empezando por las más sencillas y luego las más complejas. Esto se denomina "embudo de preguntas".
- *De control*: se utilizan para comprobar la veracidad de las respuestas de los encuestados y normalmente lo que se hace en estos casos es colocar la misma pregunta pero redactada de forma distinta en lugares separados una de la otra.
- *Amortiguadoras*: se refieren a que cuando estamos preguntando temas escabrosos o pensamos que serán reticentes a contestar, hay que preguntar suavizando la pregunta y no preguntar de modo brusco y directo.

- *De introducción o de contacto*: son las que se hacen para iniciar el cuestionario o para pasar de un tema a otro al objeto de crear un clima de confianza e interés en el entrevistado.

c) *Según su contenido*:

- *Identificación*: sitúan las condiciones en la estructura social. Ej. Edad, sexo, profesión.
- *Acción*: tratan sobre las acciones de los entrevistados. Ej. ¿Va al cine?¿fuma?.
- *Intención:* indagan sobre las intenciones de los encuestados. Ej. ¿Va a votar?
- *Opinión*: tratan sobre la opinión encuestados sobre determinados temas. Ej. ¿Qué piensa sobre...?
- *Información*: analizan el grado de conocimiento de los encuestados sobre determinados temas.
- *Motivos*: tratan de saber el por qué de determinadas opiniones o actos.

4.1.3. Estructura del cuestionario

Al redactar las preguntas hay que tener también especial cuidado en el orden en que se incluyen en el cuestionario, ya que la secuencia de las preguntas puede influir en la naturaleza de las respuestas del encuestado y, por consiguiente, sesgarlas. Para ello existen una serie de pautas generales que pueden ser muy útiles en la estructuración del cuestionario:

- Al inicio del cuestionario se debe incluir una presentación solicitando la cooperación del encuestado y especificando los objetivos de la investigación, quién la realiza y una declaración explícita de que la información que se facilita tendrá un tratamiento global, ya que la garantía de anonimato de un cuestionario es fundamental para crear una buena disposición a contestar.
- La primera pregunta debe ser de carácter general, sencillo y potenciadora del interés del encuestado. En ocasiones esta pregunta de carácter introductorio no se relaciona con las necesidades de información de la investigación, ya que su único objetivo es lograr la cooperación del encuestado y establecer una relación de armonía con él.

- Las preguntas de tipo general deben preceder a las preguntas más específicas.
- Generalmente las preguntas más sencillas deben ir al principio del cuestionario, reservando las más comprometidas para el final y dejar el espacio intermedio para las más substantivas e importantes de la investigación. Es decir, las preguntas menos problemáticas deben formularse inicialmente y de forma gradual se introducirán las preguntas más complejas y personales.
- El flujo en el proceso de las preguntas debe ser lógico de acuerdo con la perspectiva del encuestado y, al mismo tiempo, las preguntas deben agruparse en función de su temática para evitar el desconcierto del entrevistado haciendo preguntas relativas a un mismo tema en diferentes fases de la entrevista.
- Los datos de identificación del encuestado, utilizados como variables de análisis y a nivel global para comprobar la fiabilidad de la muestra deben formularse al final del cuestionario.

4.1.4. La codificación en el cuestionario

La finalidad de la codificación es facilitar la operación de contar y analizar las respuestas dadas por la muestra que ha sido encuestada. Un cuestionario bien codificado debe contener la información exacta con respecto a las variables. Por tanto, la codificación permite saber, una vez recogidos los datos, el número total de variables a analizar y su ubicación exacta en la matriz de datos. Por consiguiente, la formalización del cuestionario es fundamental para garantizar la calidad de las respuestas y del proceso de codificación. En los cuestionarios que rellena directamente el encuestado (no hay entrevista) hay que adoptar un formato claro que facilite su cumplimentación al máximo. En estos casos es mejor no complicar la respuesta haciendo que el encuestado codifique directamente cada respuesta en una casilla aparte; poner una cruz al lado de la opción seleccionada para las preguntas cerradas y dejar unas líneas para contestar es la mejor solución. Un cuestionario complejo de responder, que requiere muchas explicaciones para su cumplimentación, repercute negativamente sobre el nivel de respuesta.

Respecto a los métodos de codificación señalamos dos alternativas, en función de si las preguntas son cerradas o abiertas. Para las preguntas cerradas hemos de asignar un valor numérico a cada posible categoría de respuesta. Cada pregunta o variable ocupa un espacio físico (columnas) en la matriz de datos. En el caso de las preguntas abiertas, hay que hacer la codificación después del trabajo de campo a través de un análisis de contenido. No obstante, hay que prever el espacio físico que ocupa la variable en la matriz de datos (generalmente dos columnas y, por tanto, más de 9 categorías posibles de respuesta). *La matriz de datos viene determinada por: N sujetos (filas) × P variables (columnas).*

4.1.5. Estudio de las características técnicas

Al diseñar un cuestionario hemos de asegurarnos la máxima exactitud de los datos. Ello dependerá de dos aspectos que se han de optimizar: la precisión o fiabilidad y la validez de la información.

La validez, definida como la ausencia de sesgos, representa la relación entre lo que medimos y aquello que realmente queremos medir. Existen tres métodos para garantizarnos dicha evidencia. Los señalamos a continuación:

- El método relacionado con el **contenido** se centra en la muestra de las preguntas de un cuestionario (establecer el grado de representatividad de los contenidos recogidos en las preguntas del cuestionario) y lo que representa el completo dominio del contenido de las preguntas, por parte de los usuarios encuestados.
- El método relacionado con los criterios se centra en las relaciones estadísticas existentes entre las mediciones, para poder saber si los cuestionarios pronostican lo que deben pronosticar.
- El método relacionado con los factores está compuesto por los dos métodos anteriores y es un método guiado por la teoría, pues especifica aquello con que la medición debe o no relacionarse.

La precisión, definida como la ausencia de error aleatorio, representa la influencia del azar en nuestra medida; es decir, es el grado en el que las mediciones están libres de la desviación producida por los errores causales. Además, la precisión de una medida es lo que asegura su repetitividad (si la

repetimos, siempre da el mismo resultado). Existen diversos factores que afectan a la fiabilidad de los cuestionarios. Dos de ellos son el número de preguntas de los cuestionarios (debemos asegurarnos que contamos con un cuestionario que tiene múltiples preguntas que traten cada una de las categorías que estamos midiendo), y la muestra de usuarios sobre la que se calcula la estimación de la fiabilidad.

4.1.6. Comprobación y prueba del cuestionario. Pretest

Cuando tenemos ya diseñado el cuestionario con todas las garantías, estamos en condiciones de comenzar a encuestar a la población. Pero, nadie nos asegura que el cuestionario funcione en la práctica. No debemos olvidar que el cuestionario se ha diseñado en un despacho y no se ha demostrado su eficacia en condiciones reales. Por ello, antes de iniciar el trabajo de campo, es imprescindible probar el cuestionario sobre un pequeño grupo de población. Esta *prueba piloto* ha de garantizar las mismas condiciones de realización que el trabajo de campo real. Su misión radica en contrastar hasta qué punto funciona el cuestionario como se pretendía en un primer momento, y verificar si las preguntas provocan el tipo de reacción deseada. Por ello, sí es necesario, se han de eliminar ambigüedades y preguntas superfluas, se pueden añadir al cuestionario preguntas relevantes o cambiar el orden de éstas para agilizar el flujo de respuestas. Es frecuente hacer una primera prueba piloto con algunas preguntas abiertas para tratar de obtener las respuestas más frecuentes, y construir preguntas cerradas en el cuestionario definitivo. En este caso, más que de una prueba piloto hablaremos de un *sondeo previo*. Este tipo de técnica es muy recomendable en los casos en los que hay un gran desconocimiento de las posibles respuestas, para intentar cerrar la mayor cantidad posible de preguntas del cuestionario. La prueba piloto habrá de valorar los siguientes aspectos:

- Verificación si el cuestionario responde a los objetivos del estudio.
- Comprobación si el cuestionario es fluido. Es decir, el cuestionario debe tener lógica y consistencia interna.
- Comprensión de las preguntas y aceptabilidad por parte del encuestado.
- Idoneidad en la secuencia de las preguntas.

- Idoneidad de las respuestas cerradas preestablecidas.
- Discriminación de las preguntas.
- Valoración de los casos en que nos encontremos que los encuestados no responden al cuestionario.
- Idoneidad de todos los aspectos del protocolo de procedimientos.
- Proceso de codificación en cuestionarios autocodificables.
- Idoneidad y nivel de preparación de los encuestadores.
- Aspectos logísticos: disponibilidad, recogida y entrada de cuestionarios, la propia supervisión, etc.

En esta prueba preliminar es aconsejable utilizar encuestadores experimentados para detectar posibles preguntas que originen confusión en el encuestado, palabras que no se entiendan, problemas que se pueden presentar en las categorías normalizadas de las preguntas cerradas, etc.. Asimismo, el formato de respuesta abierta puede utilizarse en este pre-test para determinar las categorías de respuestas más apropiadas o frecuentes y, posteriormente, convertirla en una pregunta cerrada.

Algunas de las variables que nos determinan un mal diseño del cuestionario son el número de negativas a contestar preguntas, una alta proporción de "no sabe/ no contesta" y también preguntas en blanco. A no ser que se haya seleccionado erróneamente la muestra y, por ejemplo, se pretenda investigar un tema de alta cualificación entre personas con poca formación, no es normal que el porcentaje de preguntas en blanco o que no se saben contestar supere el 10% si el cuestionario está bien diseñado. Cuando ocurren estos casos, habrá que investigar las causas de los problemas, corregir los errores detectados hasta llegar a la redacción del cuestionario definitivo.

En relación al número de personas que constituyen la sub-muestra de esta prueba preliminar depende de las características de la muestra, aunque normalmente suele oscilar entre 25 y 50 personas y, lógicamente, las personas que integran la sub-muestra deben coincidir en sus características con la muestra de la investigación definitiva.

4.2. Diseños muestrales

4.2.1. Selección de la muestra

El diseño de la muestra es una etapa esencial en toda investigación, que no estudia globalmente a la población objeto del análisis. Para seleccionar a los individuos que han de formar parte de una muestra podemos considerar básicamente dos grandes grupos de técnicas los métodos de muestreo aleatorio y los de muestreo empírico. Los primeros, los de ***muestreo aleatorio***, se basan en el principio de la elección de los individuos al azar. Es decir, este tipo de muestreo tiene la posibilidad de que cada elemento de la población tenga la probabilidad de ser elegido como parte de la muestra. Por consiguiente, es el único que realmente puede garantizar la representatividad de la muestra, al asegurar la ausencia de sesgos. En el ***muestreo empírico***, la selección de los elementos de la muestra se realiza a partir de los criterios fijados por el propio encuestador o investigador. De esta forma, este tipo presenta serios problemas de validez, y en él la presencia de sesgos incontrolados es un riesgo constante.

Una vez que se ha elegido el tipo de muestreo que vamos a utilizar en la investigación, es necesario determinar el tamaño de la muestra. Habremos de determinar un número suficiente de individuos para garantizarnos la fiabilidad y validez de los resultados. De lo contrario, nuestra investigación se asentaría en una muestra sesgada, no representativa y no estaríamos en condiciones de hacer inferencias. Este problema se puede resolver utilizando un tipo de muestreo adecuado.

Por último, creemos preciso señalar que no siempre es necesario hacer una muestra. En muchos casos resulta estéril porque la población objeto es pequeña y la muestra necesaria es casi tan grande como la propia población de referencia. No obstante, existen situaciones en las que es mejor trabajar sobre una muestra, como en los siguientes casos:

1) la población objeto es muy grande;
2) el periodo de estudio es muy largo;
3) a pesar, de que la población no es muy grande, nuestras necesidades de precisión no son muy exigentes;
4) los recursos son insuficientes para hacer un estudio a toda la población, siempre es mejor tener una aproximación que no tener nada.

Para una información más completa del tamaño muestral y los procedimientos de muestreo el lector puede consultar cualquier texto de estadística aplicada.

4.3. El trabajo de campo

4.3.1. Introducción

El trabajo de campo es aquella etapa del proyecto de investigación durante la cual los encuestadores se ponen en contacto con los encuestados, administran los cuestionarios para recoger la información, registran los datos y devuelven los cuestionarios para su posterior control y procesamiento.

La planificación del trabajo de campo es un aspecto básico en el proceso de una investigación de encuestas y su desarrollo depende del método de obtención de la información que se utilice; así el trabajo de campo para un estudio basado en encuestas personales es totalmente distinto al de una investigación basada en encuestas postales.

4.3.2. Recomendaciones en el proceso de la entrevista

En las entrevistas personales y telefónicas, el encuestador puede ser una fuente de errores relacionados con:

a) falta de armonía entre el entrevistador y el encuestado
b) la formulación incorrecta de las preguntas
c) errores en el registro de las respuestas
d) falseamiento de las encuestas.

A continuación analizamos brevemente los aspectos más relevantes que hacen referencia a los cuatro elementos anteriormente reseñados para obtener mejores resultados en el proceso de la entrevista.

4.3.3. Armonía entre el entrevistador y el encuestado

Durante una entrevista, la percepción del encuestado acerca del encuestador puede afectar directamente a la capacidad de este último para esta-

blecer una armonía adecuada, ya que los entrevistadores que pueden establecer relaciones efectivas con los encuestadores son capaces de recoger información más completa y exacta.

En el caso de la entrevista personal, la forma de vestir del entrevistador y su compostura son importantes para establecer una buena armonía, ya que la mayoría de las personas los consideran como indicadores de las actitudes de las personas. En este sentido, la conducta del encuestador debe ser de gran amabilidad y el encuestado debe ver al entrevistador como una persona capaz de entender sus puntos de vista.

4.3.4. Formulación de las preguntas

Se han desarrollado una serie de pautas apropiadas para, la mayoría de las situaciones de entrevista que a continuación brevemente reseñamos:

1. Estar muy familiarizado con el cuestionario. El encuestador debe estudiar el cuestionario pregunta por pregunta para evitar cometer errores en su lectura. Las preguntas se deben leer de la forma más natural posible y en un tono de conversación.
2. Formular las preguntas exactamente como aparecen escritas en el cuestionario. El entrevistador no debe realizar cambios en la formulación de las preguntas, tales como dejar de leer una parte de la pregunta, cambiar palabras o agregarlas, etc., ya que puede distorsionar el resultado de la investigación.
3. Formular las preguntas en el orden que aparecen en el cuestionario. En el diseño del cuestionario se estableció la secuencia de las preguntas de forma que exista un sentido de continuidad en el tema tratado y se eviten sesgos en la obtención de la información.
4. Formular cada una de las preguntas especificadas en el cuestionario. La respuesta del encuestado a una de las preguntas puede contestar otra de las preguntas que aparecen después en el cuestionario. En esta situación, el entrevistador no debe omitir la pregunta que parece haber sido contestada con anterioridad, ya que puede ser una pregunta de control.
5. Utilizar técnicas de indagación para lograr que el encuestado conteste la pregunta. Las preguntas han sido diseñadas para que las con-

testen todos los encuestados que forman parte de la muestra; sin embargo, en ocasiones puede ocurrir que el encuestado no comprenda o mal interprete la pregunta, que rehúse a contestarla o que de una información incompleta o ambigua. En estos casos, el entrevistador debe utilizar ciertas técnicas neutrales de indagación como son las siguientes:

a) Repetir la pregunta. Un enfoque muy efectivo consiste en repetir la pregunta exactamente como aparece escrita en el cuestionario.

b) Pausa de expectativa. Si el entrevistador permanece en silencio o hace una pausa crea una clave efectiva hará que el encuestado tome conciencia de que se espera una respuesta más completa.

c) Repetir la respuesta del encuestado. Con frecuencia se estimula a los encuestados para que hagan comentarios adicionales después de repetirles lo que han expresado. Esto puede hacerse mientras el entrevistador registra la respuesta en el cuestionario.

d) Tranquilizar al encuestado. Si el encuestado duda al responder, puede ser de gran utilidad realizar comentarios neutrales como los siguientes: "no existen respuestas correctas o incorrectas, sólo queremos sus ideas sobre el tema" o "solamente estamos tratando de conocer las ideas de las personas acerca de este tema".

e) Preguntas o comentarios neutrales. Cuando la respuesta del encuestado es un tanto incompleta se puede recurrir a preguntas neutrales, tales como "¿algo más?" "¿alguna otra razón?" "¿podría decirme algo más de lo que piensa sobre esto?" "¿qué quiere decir?".

f) Solicitar una clarificación más detallada. Al formular la pregunta "no estoy muy seguro de lo que usted quiere decir con eso ¿podría explicármelo un poco mejor?", el encuestado puede sentir el deseo de cooperar con el encuestador y dar una información más detallada de su respuesta.

4.3.5. Registro de respuesta

El encuestador debe registrar las respuestas aportadas por el encuestado de una manera imparcial para que se pueda interpretar con precisión por

parte de los codificadores. Para ello es necesario que registre la respuesta en el mismo momento que el encuestado la expresa, utilizar las mismas palabras del encuestado, no resumir o parafrasear las respuestas del encuestado, incluir comentarios entre paréntesis, mantener el interés del encuestado repitiendo la respuesta cuando se escribe, emplear letra legible y en caso de preguntas en blanco, explicar los motivos de no contestar la pregunta.

4.3.6. Falseamiento de las encuestas

Para evitar que puedan producirse falseamientos de encuestas por parte de los encuestadores es necesario introducir unos mecanismos de control de los encuestadores que nos permiten garantizar la bondad del trabajo de campo realizado. En el siguiente epígrafe analizaremos los sistemas de control más comúnmente utilizados para asegurar que la información obtenida no ha sido falseada.

4.3.7. Control del trabajo de campo

Teniendo presente que la base de toda investigación descansa en la veracidad de la información recogida, es necesario controlar que los datos sean ciertos y que procedan realmente de la muestra seleccionada. Para verificar esto, en las encuestas personales es necesario un exhaustivo control a través de una revisión de una pequeña muestra de las encuestas realizadas por cada encuestador. Los sistemas de control utilizados habitualmente pueden ser directos o indirectos.

Los sistemas de control directos se basan en un control personal o telefónico realizando nuevamente la encuesta o preguntando si realmente le realizaron la encuesta en la fecha indicada. El porcentaje de encuestas que se recomienda controlar personal o telefónicamente oscila entre un 10% o 15% , dependiendo de la confianza que se tenga en los encuestadores.

Los sistemas de control indirectos consisten en un análisis interno y comparativo de los cuestionarios cumplimentados por cada entrevistador. Entre los métodos más utilizados podemos destacar los siguientes:

a) **El control de curvas de estabilidad.** Se establece a través de la tabulación de los cuestionarios de un entrevistador en orden secuen-

cial de realización de su trabajo. Un número elevado de respuestas en el mismo sentido para alguna pregunta, superior a las dadas por otros entrevistadores, supone un control más estricto de su trabajo.

b) **El control de consistencia de preguntas.** Se comprueba si existen las mismas respuestas o ausencia de las mismas en los cuestionarios de los encuestadores.

c) **El control de batería de preguntas.** Se comprueba a través de las preguntas diseñadas a tal efecto para controlar el grado de aceptación que se debe dar a la información obtenida.

4.4. El análisis estadístico

4.4.1. Tratamiento y análisis de la información

Se han descrito las distintas técnicas de obtención de información. Ahora bien, es necesario analizar esta información, generalmente abundante, y convertirla en conclusiones. Presentamos, pues, una serie de etapas a seguir a la hora de codificar, analizar, tratar e interpretar la información recogida a través de los cuestionarios:

- **Revisión de los cuestionarios**. En esta fase se trata de identificar y corregir las posibles fuentes de error. Para ello se revisan los cuestionarios buscando ambigüedades, como respuestas no legibles o en las que no se sabe qué cuadro se ha marcado, omisiones o incoherencias.

- **Codificación, clasificación, recopilación de cuestionarios y preparación de la base de datos.** Una vez depurados los cuestionarios se procede a la codificación de las preguntas para posibilitar el tratamiento informático. La codificación tiene por objeto sistematizar y simplificar la información procedente de los cuestionarios. En otras palabras, consiste en el establecimiento de grupos que permitan clasificar las respuestas. Conviene recordar que la codificación ha de ser lo más simple e intuitiva posible, siempre claramente unívoca, preferentemente numérica y lo más estandarizada posible. Con los códigos de respuesta asignados a cada variable, construiremos la base de datos que reflejará toda la información obtenida.

- **Análisis de las preguntas**. Una vez organizados los datos en un fichero, comienza el análisis propiamente dicho. El primer paso es estudiar cada pregunta aislada, luego las preguntas por subgrupos y las relaciones entre pares de preguntas, y por último, se estudiarán las relaciones entre todas las preguntas.
- **Conocer la evolución del estudio**. Como la percepción de los problemas cambia pareja a cómo evoluciona la satisfacción de los usuarios de un servicio de información-documentación, hay que establecer un sistema de vigilancia de los principales parámetros de percepción, que nos avise cuando haya un nuevo problema sentido, cuando las cosas no van bien desde el punto de vista de los usuarios, etc. En consecuencia, es necesario desarrollar sondeos y estimaciones durante distintos momentos del análisis.
- **Anticipación del estudio de predicción.** No es suficiente con conocer la evolución del estudio o análisis, hemos de saber además hacia dónde nos lleva. Para ello, hemos de hacer proyecciones de futuro sobre los problemas, las necesidades, las demandas y la satisfacción de los usuarios objeto del estudio. El análisis ha de permitir asociar probabilidades y formular predicciones a corto plazo. En definitiva, ha de permitir anticiparnos y modificar nuestras actuaciones.El lector interesado en profundizar en los aspectos tratados en este apartado puede consultar entre otros, nuestro texto: "Estadística e Informática (SPSS) en la Investigación Descriptiva e Inferencial" (UNED. Madrid, 2006).

4.5. El informe de resultados

La última etapa de una investigación la constituye la redacción final de informe, es decir, un documento escrito que incluye las diferentes etapas cubiertas durante el proceso de la investigación y se exponen los resultados y conclusiones extraídas del análisis aportando los datos, análisis y resultados de manera organizada y coherente.

La información que se presente ha de ser muy clara, incluyéndose únicamente aquella que sea relevante para las decisiones que hay que tomar. A continuación, exponemos algunas recomendaciones útiles que pueden ayudar en la presentación final:

- Proporcionar la menor cantidad posible de datos. Eso sí, los que expongamos en el análisis deben de ser los más relevantes.
- Hacer los comentarios e interpretaciones al final de cada grupo de resultados, diferenciando muy claramente lo que es un resultado de lo que es una interpretación personal.
- Utilizar un lenguaje sencillo, sin posibilidad de mal interpretaciones.
- La información sobre la metodología y los criterios de análisis para el estudio han de estar bien diferenciados del resto de la información.
- En general, no se deben desagregar resultados, excepto que se haya demostrado una relación significativa con la variable de estratificación.
- En estudios efectuados sobre muestras representativas hay que proporcionar las estimaciones por intervalo.
- No se han de proporcionar fracciones decimales irrelevantes. En la mayoría de los casos con un decimal es más que suficiente.
- No debemos dar información redundante.
- Siempre que sea posible, se han de proporcionar los datos en forma gráfica. Además, en todas las tablas y gráficos presentados han de incluirse los nombres de las variables y las unidades de medida.
- Al final del estudio se ha de incluir un breve resumen de los resultados y de las conclusiones. En algunos casos es conveniente platear recomendaciones operativas.
- Si existen niveles diferentes de usuarios, se debe seleccionar y segmentar la información.

Lógicamente, no existe un modelo único de informe, pues depende del tipo y profundidad de la investigación llevada a cabo, la metodología utilizada, los objetivos planteados y del destinatario del mismo. Desde este último punto de vista se pueden distinguir dos tipos de informes:

a) informes técnicos, que son los destinados a personas con alta formación técnica, para los cuales cobran singular relevancia los detalles relativos a la metodología utilizada

b) informes divulgativos que están destinados a una audiencia más general, muchas veces poco familiarizada con detalles técnicos, por

lo que en este tipo de informes los detalles técnicos tienen mucho menos interés que las conclusiones y recomendaciones que contenga, pues se supone que sobre esta información van a basarse las decisiones.

En los informes deben combinarse ambos enfoques, dando mayor importancia al aspecto divulgativo cuando los resultados del informe van destinados a directivos de empresa.

Si bien no existe un formato único y específico que defina la estructura de un informe, en términos generales podemos establecer una estructura común como la más frecuentemente utilizada:

a) **Portada**. La portada debe contener un título que resuma la esencia del estudio, la fecha, el nombre de quien presenta el informe y el nombre de la organización a quien va dirigido el informe.

b) **Índice de Contenidos**. Al objeto de ayudar a los lectores a encontrar secciones específicas que puedan ser de mayor interés, el índice enumera en forma secuencial los temas que se encuentran en el informe, junto con referencias de las páginas. En algunas ocasiones, también se utiliza un índice de tablas (figuras, ilustraciones, gráficos) para enumerar los títulos y números de página de todas las ayudas visuales.

c) **Introducción**. El objetivo de la introducciones proporcionar al lector la información básica (antecedentes) necesaria para entender el resto del informe. La naturaleza de la introducción está condicionada por la diversidad de la audiencia y su familiarización con el proyecto de investigación. Cuanto más diversa sea la audiencia, más extensa será la introducción. En términos generales, en la introducción se debe explicar claramente la naturaleza del problema y los objetivos de la investigación.

d) **Metodología**. En los apartados que integran esta sección se detallan las fuentes de información utilizadas, los aspectos relacionados con el diseño muestra y el método utilizado en la obtención de información. En definitiva, la sección de metodología debe: 1) resumir los aspectos técnicos del proyecto de investigación en un estilo comprensible y 2) permitir desarrollar una confianza en la calidad de los procedimientos utilizados.

e) **Resultados**. El cuerpo del informe está compuesto por los resultados de la investigación, los cuales deben estructurarse en función de los objetivos que se persiguen. Es decir, los resultados deben estructurarse en un flujo lógico de información y no debemos limitarnos a ofrecer una serie de frecuencias de respuesta, una para cada pregunta del cuestionario. Ello reflejaría una falta de creatividad y empeño. Con frecuencia, en la presentación de resultados se emplean tablas, gráficos e ilustraciones para explicar las diferentes relaciones entre los diversos datos analizados.

f) **Conclusiones y Recomendaciones**. Las conclusiones tratan de resumir los principales resultados que merecen especial atención y sobre la base de estas conclusiones hacer las recomendaciones más pertinentes.

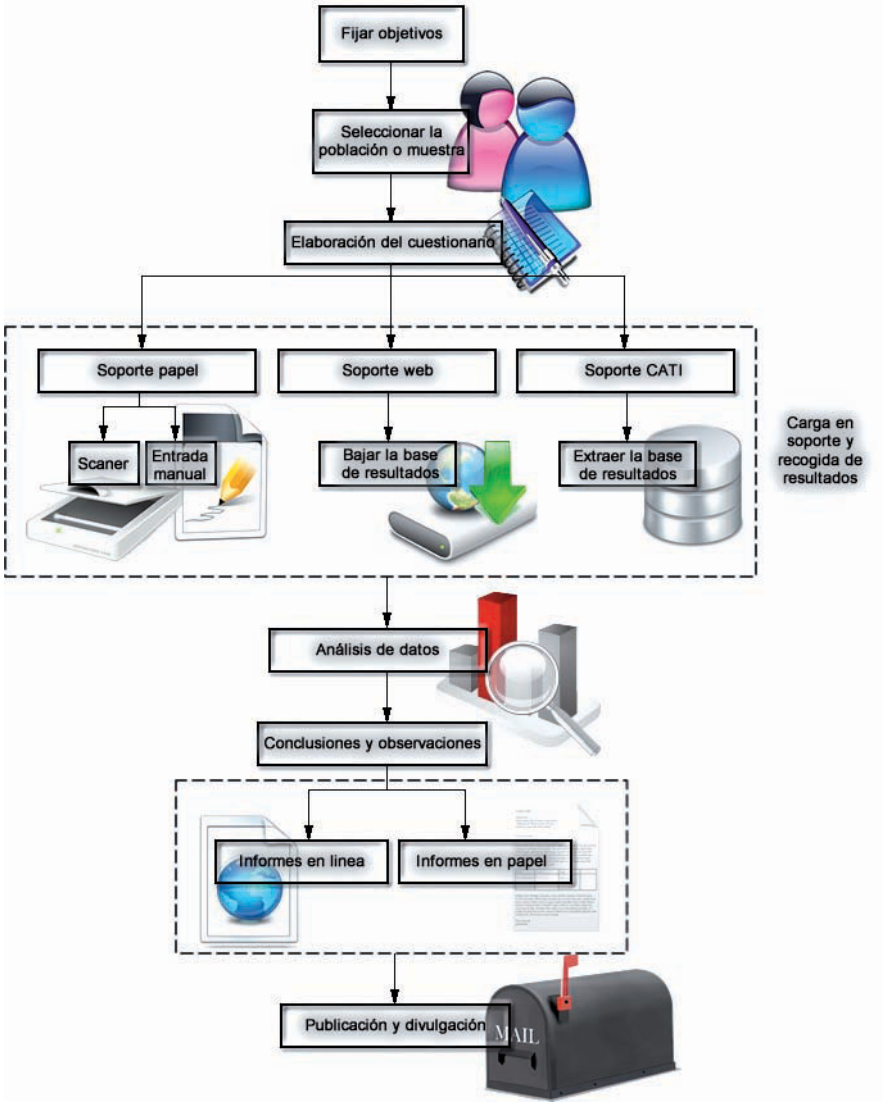
g) **Anexos o Apéndices**. Se incluyen los materiales de apoyo que no son absolutamente esenciales en el cuerpo del informe: Copia de cuestionario o formato de recogida de información, cálculos del plan de muestreo, cálculos estadísticos, que no deben incluirse en el cuerpo del informe, tablas estadísticas que no son relevantes, etc.

h) **Bibliografía**. Debe incluir todas las publicaciones o fuentes consultadas en la realización de la investigación

4.6. Ayudas informáticas para el tratamiento de encuestas

4.6.1. Programas

Existen numerosos programas para crear, publicar y administrar encuestas, de algunos se ha hecho referencia en el apartado dedicado a las encuestas on-line. En este punto, para que el lector tenga una visión de conjunto, se muestra a continuación el diagrama de bloques de un sistema integrado de tratamiento de encuestas:



4.7. Conclusiones

Resumiremos una serie puntos sobre los aspectos más relevantes a considerar a la hora poner en práctica una metodología de encuesta, de una forma coherente y rigurosa:

- En primer lugar, se han de formular los fines y los objetivos de la investigación. En los estudios inferenciales se han de plantear también previamente las hipótesis. Además, han de conocerse cuáles son las variables a considerar en el análisis.
- El cuestionario debe ser breve y concreto, ajustado a la cantidad de información que se necesita, y accesible para el entrevistado. Asimismo, hemos de intentar evitar la ocasión de que éste pueda sesgar o falsear las respuestas.
- Una pregunta válida debe parecer importante, ser concisa e inequívoca, contener sólo un pensamiento, y no presentar dobles negativas. De esta forma, se obtendrán respuestas claras y sencillas y se dispondrá de un cuestionario fácil de leer y completar.
- Por lo que se refiere a la secuenciación de preguntas, es aconsejable comenzar el cuestionario con preguntas iniciales no conflictivas que susciten interés al encuestado. Seguir con preguntas abiertas con respuestas que recojan la espontaneidad de los usuarios. Después vendrá un grupo de preguntas cerradas. Concluiremos el cuestionario con una serie de preguntas finales más conflictivas y rutinarias.
- Debe medirse en su justa medida la extensión del cuestionario, de modo que obtengamos la información buscada, pero sin cansar al entrevistado.
- Debe de elegirse la manera de llevar a cabo la encuesta: por correo, telefónica, con entrevista pero sin cuestionario, entrevista con cuestionario, con cuestionario en el mismo lugar donde se presta el servicio, etc.
- Hay que tener claros los criterios de codificación.
- En la fase de diseño, hemos de tener en cuenta si vamos a estudiar a toda la población o si por el contrario estudiaremos tan sólo una parte (una muestra) que la represente. Si se elige la segunda opción, hemos de seleccionar el tipo de muestreo a emplear, y definir el tamaño y el nivel de error.
- Al diseñar un cuestionario, hemos de asegurarnos la validez y fiabilidad de los resultados, y antes de iniciar el trabajo de campo, con el fin de mejorar la efectividad del cuestionario, hemos de someterlo a algún tipo de filtro o depuración (simulaciones de laboratorio, pre-test, pruebas piloto, etc.).

5. UN EJEMPLO DE ENCUESTA

Como se ha visto hay varios procedimientos de encuestas, cada uno con sus ventajas e inconvenientes, por cuestión didáctica vamos a presentar un cuestionario realizado con la idea de medir el clima laboral, constructo construido con la idea de tomar información de la satisfacción en el trabajo de un conjunto de empleados de una institución o empresa.

El proceso para realizar la investigación es el siguiente:

1) Fijar los objetivos

El objetivo general de la investigación es apreciar la SATISFACCIÓN del personal (Clima laboral) fijando las características de ciertos factores que lo configuran.

El clima laboral se puede definir como: un conjunto de características objetivas y relativamente permanentes de la Organización, percibidas por los individuos pertenecientes a la misma, que sirve para dotarla de una cierta personalidad y que influye en el comportamiento y las actitudes de sus miembros.

2) Muestra

Será necesario fijar previamente el universo objeto de estudio, al que naturalmente queremos generalizar los resultados obtenidos con una muestra representativa del mismo. También, por cuestiones didácticas, se supone que la población objeto es una empresa editorial de 1200 empleados.

Si se supone un muestreo aleatorio simple, y se quiere estimar la proporción que se presenta una variable relevante, al no tener datos de la misma se supone el caso más desfavorable del muestreo p=q=0,5 y suponiendo una distribución normal de la misma con un intervalo de confianza de 2σ y un error de estimación del 5% (0,05). El tamaño muestral será:

```
N<-1200 # Tamaño de la población
e<-0.05 # error de estimación
k<-2 # grado de confianza
p<-0.5
q<-0.5
D<-(e/k)^2
```

```
a<-(N/(N-1))
b<- (D/(p*q)+(1/(N-1)))
n<-round(a/b)+1
n # Tamaño de la muestra
```

Realizado el programa precedente en R, el tamaño muestral es n=301 casos.

Teniendo el listado del total de los 1200 empleados por un procedimiento aleatorio se extraerán los 301 empleados y se les pasará el cuestionario presentado a continuación.

```
sec<-sample(c(1:N),n,replace=FALSE)
sec # Presenta los empleados elegidos
```

Ejecutado el programa precedente en R se llevará a cabo la encuesta a los empleados seleccionados.

3) El cuestionario

El cuestionario del ejemplo está formado por 22 ítems de escala continua (valoración de 1 a 9) agrupados en cuatro escalas: organizativa, profesional, relacional y económica. Un bloque de 15 ítems para valoración (en escala de 1 a 9) de los aspectos más importantes y actuales de la política de recursos humanos: organización del trabajo, retribución, innovaciones, etc. Un conjunto de 10 preguntas sobre condiciones de trabajo que los encuestados deben ordenar y finalmente 8 preguntas de clasificación (sexo, edad, grupo laboral, etc.).

==========

ENCUESTA CLIMA LABORAL

ESTOY COLABORANDO CON LA UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA EN LA REALIZACIÓN DE UN ESTUDIO PARA CONOCER LA OPINIÓN DE EMPLEADOS, SOBRE DISTINTOS TEMAS. LOS JUICIOS Y DATOS QUE LE VOY A PEDIR TENDRÁN UN TRATAMIENTO EXCLUSIVAMENTE ESTADÍSTICO, DE MANERA QUE TODA LA INFORMACIÓN QUE ME DE USTED QUEDARÁ GARANTIZADA POR UN TOTAL ANONIMATO.

MUCHAS GRACIAS POR SU COLABORACIÓN

Voy a referirme a una serie de cuestiones relacionadas con su trabajo en la Empresa. Valore desde 1 (nada importante) a 9 (muy importante) la percepción actual de los siguientes aspectos de su actividad en la empresa. Indique su valoración considerando el no sabe/no contesta (NS/NC) con 0.

1. Dimensión profesional

	Valoración
1. En general, y teniendo en cuenta todos los aspectos de su trabajo, cómo valoraría su grado de satisfacción laboral?	
2. Adaptación a las exigencias del puesto de trabajo	
3. Autonomía en el trabajo	
4. Conocimientos de los objetivos de la empresa	
5. Su realización personal	
6. La comunicación interna dentro de la Empresa	
7. Capacidad de promoción dentro de la Organización	
8. La capacidad de maniobra en su carrera profesional	
9. Interés por su formación dentro de la Empresa	

2. Dimensión económica

	Valoración
10. El salario percibido	
11. El modo de distribución de la remuneración	
12. Los beneficios sociales y otros conceptos retributivos	

3. Dimensión relacional

	Valoración
13. Relación con el jefe inmediato	
14. Respeto a los usos y costumbres de la Organización	
15. Sentido de autocrítica en su relación con los demás	
16. Relación con los compañeros (colegas)	
17. El reconocimiento y consideración social que tiene el puesto que ocupa	

4. Dimensión organizativa

	Valoración
18. La cultura empresarial de su Organización	
19. La imagen externa de la Empresa	
20. La organización de la Empresa	
21. La satisfacción con su labor diaria	
22. El sentimiento de pertenencia a su Empresa	

A continuación le voy a citar una serie de instituciones y aspectos que se relacionan con la Empresa. Valore desde 1 (nada importante) a 9 (muy importante) su funcionamiento en la Empresa. Indique su valoración considerando el no sabe/no contesta (NS/NC) con 0.

	Valoración
23. Innovaciones con ensayos de nuevas ideas y técnicas	
24. Dirección	
25. Mandos intermedios	
26. Organización del trabajo dentro de la Empresa	
27. Formación	
28. Selección de personal	
29. Ayuda de vacaciones	
30. Economatos o ayuda en concepto similar	
31. Servicios médicos o mutuas médicas	
32. Horario y jornada laboral	
33. Fondos de pensiones u otras mutuas de previsión	
34. Traslados y/o cambios de acoplamiento	
35. Retribución fija	
36. Retribución variable (incentivos)	
37. Representación sindical	

Vamos a considerar ahora algunas cuestiones relacionadas con las condiciones generales de trabajo en su Empresa.

En las preguntas que le voy a mostrar aparecen diez factores relacionados con las condiciones de trabajo. Ordénelos por favor, según la importancia que tienen para usted.

	Ordenación
38. Sueldo o incentivos	
39. Relación con mis compañeros	
40. Formación	
41. Planes de promoción y carrera	
42. Ambiente laboral	
43. Prestigio o reconocimiento social	
44. Planes Sociales	
45. Estabilidad en el empleo	
46. Jornada laboral y horario	
47. Vacaciones	

Variables de clasificación

48. SEXO Hombre ..... 1 Mujer 2	52. SECTOR DE ACTIVIDAD Industria ..... 1 Comercio..... 2 Servicio 3
49. EDAD 30 años ó menos ..... 1 31-40 años..... 2 41-50 años..... 3 51 ó más años 4	53. TITULARIDAD Pública ..... 1 Privada 2
50. GRUPO LABORAL Trabajadores manuales ..... 1 Trabajadores administrativos ..... 2 Técnicos ..... 3 Mandos..... 4	54. ESTUDIOS TERMINADOS Estudios Primarios..... 1 Bachiller Elemental (EGB), Oficialía Industrial, FP1..... 2 BUP, Maestría Industrial, FP2..... 3 Estudios Universitarios Medios (Ingenieros Técnicos, Diplomados, etc.)..... 4 Estudios Universitarios Superiores..... 5

51. TAMAÑO DE LA ENTIDAD Pequeña (menos de 100 empleados) ..... 1 Mediana (entre 101 y 500 empleados) ..... 2 Grande (más de 500 empleados) 3	55. LUGAR DE REALIZACIÓN DE LA ENTREVISTA En el puesto de trabajo, próximoal jefe inmediato ..... 1 En el puesto de trabajo, sin la proximidad del jefe inmediato..... 2 En lugar apartado (sala de formación, de reunión o de descanso, etc.)..... 3 Fuera del edificio de trabajo..... 4 Otros..... 5

==========

4) Carga en soporte y recogida de resultados

Según lo expuesto con anterioridad será necesario poner la encuesta en un soporte para que sea cumplimentada por los encuestados. En el ejemplo propuesto el soporte fue papel y se realizó la lectura de datos y creación de la base de resultados mediante procedimiento manuales, es decir, mediante la introducción de los mismos en una base de SPSS.

5) Análisis

El análisis de los datos de la encuesta se realizará en diferentes etapas: un estudio descriptivo de los distintos factores recogidos en el cuestionario, un estudio relacional con respecto a las variables de clasificación y una última etapa de modelización, donde se perfilará el colectivo de empleados con relación a los factores. Con este último análisis daremos una visión globalizada de la Empresa aportando una caracterización del Clima Laboral de la misma.

6) Conclusiones y observaciones

Como fruto de los análisis estadísticos anteriores se elaborará unas conclusiones, se realizará las observaciones que se estimen pertinentes y se comentarán las limitaciones que pueda tener la investigación.

El lector interesado en profundizar en los análisis estadísticos, conclusiones y observaciones de un estudio realizado con el cuestionario de clima laboral puede consultar la comunicación del autor presentada en el XII Congreso de Pedagogía (2000, Tomo II, pág. 516).

7) Publicación y divulgación de resultados

Cualquier estudio o investigación se realiza para que otros investigadores conozcan los resultados y se pueda generalizar las conclusiones a un colectivo mayor. En este ejemplo la divulgación se realizó en el marco del XII Congreso de Pedagogía.

BIBLIOGRAFÍA

Aparicio Pérez, Félix (1991). *Tratamiento informático de encuestas*. Ra-Ma. Madrid.

Bosch, J. y Lluís, C. (1993). *Encuestas telefónicas y por correo*. Centro de Investigaciones Sociológicas. Madrid.

Cohen, Louis (1990). *Métodos de investigación educativa.* La Muralla. Madrid.

Díaz de Rada Igúzquiza, Vidal (2002). *Tipos de encuestas y diseños de investigación.* Universidad Pública de Navarra, D.L. Pamplona.

Gil Pascual, J. A. (2008) 2.ª reimp. *Bases metodológicas de la investigación educativa* (Análisis de datos). UNED. Madrid.

Gil Pascual, J. A. (2006) 2.ª ed.. *Estadística e Informática (SPSS) en la Investigación descriptiva e inferencial*. UNED. Madrid.

Gil Pascual, J. A. (2000). El diagnóstico y la orientación de los recursos humanos en las organizaciones: estudio de Clima Laboral en la Empresa, Actas del XII Congreso Nacional y I Iberoamericano de Pedagogía. Tomo II. Resúmenes de Comunicaciones. Madrid, septiembre 2000, pp. 516-517.

Groves, R. M. (1989). *Survey errors and survey costs*. New York, Wiley.

Hyman, Herbert (1984). *Diseño y análisis de las encuestas sociales*. Amorrortu. Buenos Aires.

Kish, Leslie (1972). *Muestreo de encuestas.* Trillas. México.

Kish, Leslie (1995). *Diseño estadístico para la investigación*. Siglo XXI. Madrid.

Manzano, V. G.; Rojas, A. J. y Fernández, J. S. (1996). *Manual para encuestadores: fundamentos del trabajo de campo, aspectos prácticos*. Ariel. Barcelona.

Miguel, Amando de (1997). *Manual del perfecto sociólogo.* Espasa-Calpe. Madrid.

Rojas Tejada, Antonio J.; Fernández Prados, Juan Sebastían (1998). *Investigar mediante encuestas: fundamentos teóricos y aspectos prácticos*. Síntesis. Madrid.

Santesmases Mestre, Miguel (2001). *Dyane: diseño y análisis de encuestas en investigación social y de mercados*. Pirámide. Madrid.

Santos Peñas, Julián... [et al.] (2003). *Diseño de encuestas para estudios de mercado:técnicas de muestreo y análisis multivariante*. Centro de Estudios Ramón Areces. Madrid.

Summers, G. F. (1976). *Medición de actitudes*. Trillas. México.

Capítulo 6
LA ENTREVISTA

1. Introducción
2. Características de la entrevista
3. Fases de la entrevista
4. Tipos de entrevistas
   4.1. La entrevista estructurada
   4.2. La entrevista no estructurada
   4.3. La entrevista de grupo o grupal
5. Análisis de datos
   5.1. Tamaño de la muestra
6. Un ejemplo de entrevista

Objetivos

- Valorar la utilidad de la entrevista en las investigaciones sociales
- Conocer las distintas variantes de entrevista y sus limitaciones
- Saber elegir el modelo de entrevista más apropiado a un contexto concreto
- Diseñar una entrevista
- Realizar los controles técnicos apropiados para controlar la calidad de la entrevista
- Saber implementar una entrevista
- Recoger y analizar los datos recogidos de una entrevista

1. INTRODUCCIÓN

La entrevista es una conversación verbal entre dos o más personas con una finalidad. Como instrumento, es un proceso de recogida de información con dos componentes el entrevistado y el entrevistador.

Son muchos los fines de una entrevista: puede usarse como un medio de evaluación o valoración de las personas en los procesos de selección de personal en una empresa; por motivos terapéuticos, como puede ser una entrevista psiquiátrica; para reunir datos y en general en los procesos de investigación. En este capítulo nos centraremos en esta última aplicación, en la llamada "entrevista de investigación".

Así, por ejemplo, la entrevista del profesor con un estudiante particular, constituye un valioso recurso para obtener información sobre los rasgos de su personalidad.

Generalmente, una entrevista se suele conformar con una serie de preguntas que se plantean al entrevistado pensando en un objetivo concreto, cuyo fin es conocer sus puntos de vista sobre ciertos aspectos de un tema; para ello se preparan de antemano una serie de preguntas que, dentro de un carácter más o menos informal, se hacen en el desarrollo de la entrevista.

La entrevista maestro-alumno difiere un poco del planteamiento general de las mismas; en este tipo de entrevistas, interesa sobre todo recoger información sobre las reacciones emocionales que presenta y el grado de confianza que manifiesta hacia el educador. Necesita realizarse en un clima que favorezca la comunicación, sin demasiados formalismos con un tono conversacional en la que el maestro no deberá forzar de ninguna manera el grado de intimidad de las respuestas. A través de esta entrevista se puede obtener información importante acerca de: la situación familiar del estudiante, sus sensaciones en el grupo, sus dificultades en el estudio, las actividades realizadas, etc.

La entrevista, como técnica diferenciada de investigación, principalmente tiene tres fines:

a) Recogida de información

b) Probar hipótesis o sugerir otras nuevas

c) Usarse en conjunción con otras técnicas

El término recogida de información es muy amplio y en la interacción entre personas se concreta en saber del entrevistado sus gustos, sus valores, sus creencias, sus actitudes, etc.

La entrevista sirve para identificar las variables a la hora de probar hipótesis e incluso, con la información del entrevistado, plantearse hipótesis no pensadas por el investigador.

Al hablarse de la utilización de la entrevista en conjunción con otras técnicas, se está pensando en el apoyo en la investigación para dar respuesta a interrogantes que otras técnicas, como la encuesta, no ha podido resolver por lo encorsetado de su procedimiento.

2. CARACTERÍSTICAS DE LA ENTREVISTA

La entrevista como técnica de investigación es considerada como una de las formas de los métodos de encuesta. Es un método que comprende la reunión de datos a través de una interacción oral directa entre individuos. Permite una mayor profundidad que el cuestionario.

La entrevista está mucho más estructurada que la observación directa. En ella las respuestas son abiertas y el entrevistador ha de ser una persona preparada. La dificultad estriba en el análisis posterior de la respuesta.

La entrevista se caracteriza por ser:

a) Un medio potente de transferencia de información. El entrevistado tiene la capacidad de informar de todos los aspectos de su vida.

b) Un medio no acotado de información que el entrevistador debe modular para no introducir subjetividad en la misma, de ahí la conveniencia de un trabajo previo para fijar temática y formar adecuadamente al entrevistador.

c) Un encuentro que comparte muchos de los rasgos de las interacciones de la vida diaria (Cohen, 1990, p. 382).

3. FASES DE UNA ENTREVISTA

La entrevista se puede dividir en varias fases:

1) **Producción:** Comprende todos los requisitos previos para la ejecución de la entrevista:

- **Elegir al entrevistado.** Se debe escoger a una persona que resulte confluente con la temática a explorar y que pueda aportar alguna experiencia o idea valiosa. También se elegirá el tiempo y el lugar de la entrevista.
- **Recoger información sobre el o los entrevistados**. Sus vivencias, sus trayectorias profesionales, etc. e indudablemente sobre la temática de la entrevista.
- **Elaborar el cuestionario.** Es decir, las preguntas que se consideran adecuadas. Es importante agrupar y ordenar bien las preguntas, ya que el cuestionario es una valiosa guía en el momento de llevar a cabo la entrevista oral.

2) **Realización:** Para realizar la entrevista es importante cumplir con las fórmulas de cortesía como el saludo, la presentación, el agradecimiento y la despedida. Además es vital que el entrevistador guíe la conversación y que se muestre seguro de lo que habla, mediante el uso de preguntas breves, correctas y expresadas de forma clara.

3) **Registro y montado:** Para el registro propiamente dicho se puede utilizar el vídeo, casete, audio digital o simplemente las anotaciones de lápiz y papel realizadas durante la entrevista. En cualquiera de los casos se deberá contar con el permiso previo de encuestado.

En el montado o grabación de la entrevista se hace una selección, un recorte y una confección de la conversación sostenida. También se señalan algunos componentes no verbales que sitúen la conversación o que den cuenta del estado de ánimo del entrevistado.

Si bien se puede jugar con el orden y la distribución del material obtenido, es una exigencia ética conservar el tono general de la conversación, res-

petar el contexto en que fue dicha cada cosa y abstenerse de crear situaciones o frases que no se produjeron.

4) **Informe:** Es la fase final, en ella se da a conocer el contenido de la entrevista. Si la entrevista es el elemento primario de recogida de información su contenido vendrá reflejado de forma literal en los anexos de la investigación y servirá de base para el tratamiento estadístico de la información.

Si la entrevista se quiere publicar en algún medio de información, el entrevistador puede determinar el diseño, la extensión y la ubicación que se le da al texto y además por ética informar al entrevistado o entrevistados de tal cuestión.

En cualquiera de los casos la entrevista debe ser sometida a una evaluación del proceso y producto de la misma y a un control, mediante juicio de expertos y análisis de ítems de las preguntas que han servido de base a la misma.

4. TIPOS DE ENTREVISTAS

De forma general la entrevista puede ser: ***dirigida o estructurada*** —con guía previa de preguntas a contestar por el entrevistador— y ***no dirigida o sin estructurar,*** sin guía o libre; también puede ser **individual** o **grupal**. Un caso especial de estas últimas es el de la ***entrevista en profundidad***-, conversación entre iguales, no estructurada, dinámica, abierta; un caso particular de este tipo de entrevistas es el de las **historias de vida**.

4.1. La entrevista estructurada

Las **entrevistas estructuradas** tienen como misión sondear las opiniones de un individuo o grupo con el propósito de describirle o recoger información para confeccionar un cuestionario que permita profundizar en las características u opiniones de los colectivos. A la hora de realizar este tipo de entrevistas es importante controlar la influencia realizada por el entorno donde se desarrolle. Este entorno lo conforman, por una parte, el entrevistador, cuya pericia y rigor —sin ocultar información relevante— en la realización de la misma, es preponderante para su resultado; y por otra, el medio utilizado —personal o por teléfono—, cuya influencia está siendo un continuo foco de debate debido a la proliferación de las entrevistas telefónicas. El

orden al realizar las preguntas es muy importante. La forma de presentación y realización de las preguntas han sido anteriormente destacadas como habilidad del entrevistador. Hay que evitar dar claves reveladoras de la respuesta más aceptada o adecuada.

En la entrevista estructurada el entrevistador:

- Controla la secuencia de las preguntas siguiendo un patrón previamente establecido.
- Enuncia las preguntas siguiendo un esquema prefijado.
- No altera el orden ni la formulación de las pregunta según su esquema.
- Nunca añade preguntas nuevas.
- No expresa su opinión personal y se mantiene imparcial.
- Adopta un estilo de equilibrio en el trato pero siempre impersonal.
- Mantiene un grado de empatía normal respecto al entrevistado.

Un caso muy importante en investigación educativa es la denominada ***entrevista de investigación.*** Debe estructurarse previamente en los siguientes puntos: especificación de las variables objeto de estudio y construcción de una guía de la entrevista. En esta última se reflejarán las preguntas a realizar, indicaciones sobre acciones ante posibles respuestas, así como el modo y maneras de realizar las preguntas.

Kerlinger, citado por Cohen (1990, p. 384), propone tres tipos de ítems para la entrevista de investigación: *ítems fijos-alternativos, ítems abiertos e ítems de escala*. Los primeros son de tipo dicotómico, aunque a veces se introduce la opción no sabe.

Ejemplo:

¿Cree que el absentismo escolar es ahora mayor que hace 20 años?

Sí ..... No..... No sabe .....

Los ítems abiertos, no hay limitación en la respuesta.

Ejemplo:

¿Qué materias escolares le resultan de más fácil comprensión?

.....

.....

La escala, donde la contestación a una pregunta se conforma mediante una escala.

Ejemplo:

¿En los comedores escolares se debería incrementar la carne en los menús?

Muy de acuerdo De acuerdo Duda En desacuerdo Muy en desacuerdo

También Cohen (1990, p. 386) propone ciertos modos de preguntas y de respuestas. Las preguntas pueden:

a) Adoptar forma directa o indirecta. En el primer caso se pregunta por algo muy concreto y en el segundo por algo más general.

b) Tratar de un tema general o un tema específico.

c) Invitar a contestaciones objetivas o de opinión.

d) Hacer preguntas o emitir afirmaciones.

Las respuestas pueden presentar diversos modos:

a) Para cumplimentar espacios en blanco.

b) Tabular, es decir, mediante una tabla a rellenar.

c) Mediante una escala.

d) Para ordenar por rango.

e) De lista, donde se pide que seleccione una de las alternativas de la lista presentada.

Tuckman (1972) relaciona el modo de respuesta con el tipo de dato utilizado en los análisis estadísticos.

Tabla 6.1. Modos de respuesta y tipo de dato

Modo de respuesta	Tipo de dato	Principales ventajas	Principales inconvenientes
Para cumplimentar	Nominales	Menos parcialidad: gran flexibilidad de respuesta	Más dificultad para puntuar la respuesta
De escala	Intervalos	Fácil de puntuar	Lleva tiempo; puede haber parcialidad
De rango	Ordinales	Fácil de puntuar; fuerza de discriminación	Difícil de completar
De lista o categórica	Nominales	Fácil de puntuar; fácil de responder	Proporciona menos y menores opciones

Por otra parte, la entrevista estructurada tiene las siguientes ventajas e inconvenientes:

Tabla 6.2. Ventajas e inconvenientes de la entrevista estructurada

Ventajas	Inconvenientes
Facilita la elaboración de informes	Los entrevistados se pueden sentir molestos por la rigidez de las contestaciones
Fácil de administrar y valorar	Alto coste en preparación de la entrevista
Evaluación objetiva	La estructuración reduce la posibilidad de respuestas espontaneas y abrir líneas anexas al objetivo de investigación
Se necesita poco entrenamiento del entrevistador	El alto nivel de estructuración acota el campo de posibles aplicaciones

4.2. Entrevista no estructurada

En relación con la **entrevista no estructurada**, el ejemplo más representativo en investigación educativa lo da la ***entrevista etnográfica***.

Para Goetze y Lecompte(1988) la entrevista etnográfica se caracteriza por su carácter conversacional. Se realiza con un objetivo determinado, pero no con el deseo de contrastar una teoría, modelo o supuestos determinados como explicación de un problema. Debe realizarse de forma dirigida,

pero ganándose la confianza del entrevistado, para facilitar la transmisión de información de éste.

El entrevistado debe expresarse con naturalidad, en su propio lenguaje, sin intervención del entrevistador para resumir o matizar los términos.

La entrevista etnográfica suele comenzar hablando de temas intrascendentes para paulatinamente entrar en la temática objetivo de la misma. El entrevistador tendrá un esquema de las preguntas que quiere realizar como profundización de alguna cuestión que le preocupa, e irá desgranando las preguntas, sin presionar al entrevistado, de forma natural, sin prisas.

Otro ejemplo de entrevista no estructurada es la ***biográfica***, quizás una variante de la etnográfica. Su objetivo es profundizar en la visión que tiene una persona de su propia vida y experiencias, destacando los hechos más relevantes y el significado que los mismos han tenido para él.

En la entrevista no estructurada el entrevistador:

- Controla la secuencia de las preguntas en función de las contestaciones del entrevistado
- Enuncia las preguntas sin un esquema prefijado
- Añade preguntas si es preciso según las circunstancias de la entrevista
- Puede explicar sus sentimientos y posicionamiento
- Explica las preguntas cuando se precise
- Establece un punto de equilibrio entre familiaridad y profesionalidad
- Intenta conseguir un alto grado de empatía con el entrevistado

Respecto a las tipologías de preguntas que se pueden suscitar en una entrevista etnográfica, Spradley (1979, p. 223) y Patton (1980, p. 207 y ss) plantean las siguientes:

a) Descriptivas: Son las más fáciles y comunes en las entrevistas; con ellas se trata tanto de reconocer el lenguaje del informante como de conocer la forma particular con la que describe un acontecimiento, sus decisiones, su trabajo.

   Ejemplo:

   — Dices que te aburriste en la reunión, pero ¿qué pasó?

b) Estructurales:

Se quiere preguntar sobre cuestiones de configuración.

Ejemplo:

¿Como profesor cómo planificas los laboratorios?

c) De Contraste: Permiten conocer qué quiere decir el informante cuando emplea diversos términos durante sus explicaciones o respuestas.

Ejemplo:

— Cuando dices que el ambiente del colegio es muy bueno, ¿a qué te refieres?

d) De Opinión/Valoración: Permiten comprender qué opinan y cómo valoran los informantes acontecimientos, sucesos, acciones y decisiones, propias o ajenas.

Ejemplo:

— Algunos de tus compañeros me han dicho que la situación es bastante complicada, ¿crees que es así?

e) De Sentimientos: Con las que se anima al informante a expresar sus emociones y sus sentimientos sobre la experiencia vivida.

Ejemplo:

— ¿Te gusta salir de excursión?

f) Demográficas o de Identificación: Con este tipo de preguntas es posible identificar las características personales, sociales y profesionales de la persona entrevistada.

Ejemplo:

— Me has dicho que sois una familia numerosa: además de tus padres, ¿cuántos hermanos y hermanas tienes?

A lo largo de una entrevista el investigador puede adoptar diversas medidas para promover la locuacidad del informante, con distintos grados de intervención de forma directiva (Whyte, 1982: 112).

a) un simple movimiento con la cabeza, asintiendo, negando o mostrando interés (Inf.: Y así, el colegio se puso tranquilo; Inv.: Ahá.);

b) repetir los últimos términos del informante (Inv.: ¿Así se puso tranquilo?);

c) emplear estas últimas frases para construir una pregunta en los mismos términos (Inv.: ¿Y por qué se volvió tranquilo? (o) ¿Cuándo se puso tranquilo?);

d) formular una pregunta en términos del investigador sobre los últimos enunciados del informante (Inv.: Y ahora que está tranquilo, ¿cuál es la diferencia en el colegio comparando con otros tiempos?);

e) en base a alguna idea expresada por el informante en su exposición, pedirle que amplíe (Inv.: Ud. me decía que antes la gente era más pacífica. ¿Qué cosas pasaban entonces para que la gente fuera así?);

f) introducir un nuevo tema de conversación.

Uno de los problemas que plantea la entrevista es el registro de la misma: problemas de privacidad y veracidad se relacionan con éste. La mayoría de las entrevistas suelen grabarse en soporte de casete o audio digital, de esta forma no se distrae la línea de observación del investigador.

La entrevista no-estructurada tiene las siguientes ventajas e inconvenientes:

Tabla 6.3. Ventajas e inconvenientes de la entrevista no-estructurada

Ventajas	Inconvenientes
El entrevistador tiene mayor flexibilidad al poder realizar la pregunta que crea oportuna en cada momento	El entrevistador puede introducir sus sesgos en las contestaciones debido a la no acotación
El entrevistador puede explorar áreas que surjan sobre la marcha	Puede introducirse información extraña que genere "ruido" en el proceso de confección del informe
Puede producir información de temas que a priori se consideraron irrelevantes	El análisis e interpretación de los resultados resulta más laborioso
	Alarga el tiempo de tratamiento de los datos al tener que retirar "lo sustancial" de "superfluo"

4.3. Entrevista de grupo o grupal

Un tipo de entrevista muy utilizado últimamente en investigación es la ***entrevista de grupo***. Aunque presenta distintas variantes —Fontana y Frey (1994), tabla 6.4— todas se caracterizan porque el entrevistador debe manejar el grupo logrando que todos sus miembros participen, de ahí que éste juegue un papel a caballo entre director y moderador.

Tabla 6.4. Tipos de entrevistas de grupo y sus características (Fontana y Frey, 1994 p. 365)

Tipo	Construcción	Papel del entrevistador	Formato de las cuestiones	Propósito
Grupo de discusión	Formal- prefijada	Directivo	Estructurada	Exploratorio pretest
Brainstorming	Formal o informal	No directivo	No estructurada	Exploratorio
Nominal/Delphi	Formal	Directivo	Estructurada	Pretest exploratorio
Campo, natural	Informal, espontánea	Semi directivo	No estructurada	Exploratorio, fenomenológico
Campo, formal	Prefijada	Semi directivo	Semi estructurada	Fenomenológico

El capítulo 9 está dedicado a todo lo relacionado con la toma de datos a grupos, por tanto será en dicho capítulo donde se detallarán los distintos tipos expresados en la tabla anterior.

5. ANÁLISIS DE DATOS

Tanto la entrevista estructurada como la no estructurada, si se quiere realizar un análisis estadístico de datos necesita previamente la categorización de los mismos. En dicho proceso se deberá tener en cuenta:

- Reproducir comentarios importantes y, en las entrevistas grupales, si es posible, anotar los nombres de las personas que los han hecho.
- Organizar los comentarios por temas y subtemas tanto si se realiza con el ordenador o mediante “recortes” si el proceso es manual.
- Ordenar los temas según su importancia.
- Crear un reflejo exacto y equilibrado de lo que en realidad se dijo.

En la entrevista etnográfica el investigador trata de proporcionar una imagen fiel a la vida de lo que la gente dice y del modo en que actúa respecto al tema investigado. El investigador no trata de probar sus ideas, sino sólo de demostrar que son plausibles.

El análisis de datos en la entrevista etnográfica es un proceso en continuo, y se puede decir que pasa por tres fases: la 1ª fase es de descubrimiento: identifica temas y desarrolla conceptos y proposiciones. La 2ª es cuando los datos ya han sido recogidos, incluye la codificación de los datos y el refinamiento de la comprensión del tema de estudio. Y en la 3ª fase el investigador trata de relativizar sus descubrimientos, es decir, de comprender los datos en el contexto en que fueron recogidos.

5.1. Tamaño de la muestra

En la entrevista estructurada, dada su característica, la metodología es similar al procedimiento visto para la encuesta. En la entrevista etnográfica se siguen otros procedimientos. Como explica McCracken (1988, p. 17), en éste tipo de estudios el asunto no es el de la generalización, sino el del acceso: "el propósito de la etnográfica no es descubrir cuántos, y qué tipos de personas comparten ciertas características. Se trata de obtener acceso a las categorías culturales y los supuestos desde los cuáles una cultura construye el mundo". McCracken insiste en que la muestra de informantes en las entrevistas etnográficas no tiene como objetivo representar una parte del mundo más amplio, sino atisbar el complicado carácter, organización y lógica de la cultura.

Lo que se acostumbra, por lo tanto, es ir realizando entrevistas hasta que la diversidad de respuestas de los informantes deja de ser significativa y empiezan a repetirse los datos (cfr. Taylor y Bogdan, 1994). Dependiendo de la complejidad de los fenómenos y procesos estudiados, esto puede significar unos pocos o muchos informantes.

6. UN EJEMPLO DE ENTREVISTA

En el contexto de una investigación sobre cuidadoras, se muestra la transcripción de parte de una entrevista realizada a una cuidadora de personas mayores.

P: Bueno, Antonia, pues lo que quería es un poco que me cuentes, mi trabajo es sobre, bueno pues, como se llega a trabajar cuidando ancianos, que trabajo se desarrolla, cuando se cuidan a estas personas, que relación se tiene con la familia... Pero para un poco, para ubicarnos me gustaría que me comentases, bueno, pues desde que tu sales de tu país, de donde eres, por qué sales, que te lleva a venirte aquí España y como se va desarrollando todo hasta ahora.

R: Vale. Bueno, en primer lugar, yo tengo en España año y siete meses y... pues, salgo de Colombia en junio de hace dos años ya, que me fui por una invitación familiar a Suiza y apoyar y ayudar a una prima mía que que iba a dar a luz, entonces, este, bueno, yo es que emigro por razones familiares a España donde me contacté con con una amiga, no, que es la única que tengo en Europa y, bueno, ella me invitó aquí. Eh..., al venirme de Suiza, que estaba, este, con con visado por seis meses y al decidir venir a España pues, eh, pasado el tiempo me convierto en en en irregular, no, y bueno, pues dada esa esa consideración, tengo que buscar trabajo en instituciones de ayuda social, como Caritas, el Ayuntamiento, Cruz Roja, no, que que me puedan dar la opción de trabajar sin papeles. Es por eso que yo estoy trabajando para una persona mayor desde hace un año y anterior a esta cuando recién empecé pues estuve dos meses trabajando con otra.

P: O sea, tu sales de Colombia y te vas a Suiza ¿En qué trabajabas en Colombia?

R: Como trabajadora Social.

P: Eres trabajadora Social, trabajabas allí como Trabajadora Social y ¿Por qué fue salir de Colombia teniendo allí, tenías tu, estabas en ese momento trabajando? ¿Cómo era la situación?

R: Bueno, la situación en el país es muy difícil, económicamente Colombia es un país subdesarrollado, políticamente la economía no se maneja bien y la situación político-social pues hay mucha corrupción y esto hace de que no haya una equidad y una justicia social y tampoco un una buena política de empleo en mi país, hay mucho desempleo, hay subempleo y los trabajos pues no cuentan o no te ofrece pues ni seguros ni beneficios ni puedes este programarte a futuro pensando de cara a la familia a los hijos, no, al bienestar que tu quieres prodigar a todos y es por eso que en el anhelo de buscar la superación y el tener otras oportunidades, es que decidí aceptar la propuesta de mi prima que era temporal porque yo me iba a quedar un lapso máximo de tres a seis

meses pero que que me permitiría quizás pues tener un ahorro o impulsar algo, pero dada la situación que atravesé allá muy dura y que yo no soporte ni me adapté porque por razones familiares, personales y... bueno, no quise quedarme en Suiza, pudiéndome quedar también, si en todo caso iba a estar ilegal pues me podía quedar en Suiza.

P: Tenías esta familia.

R: Exacto. Pero pero también por otro lado tenía una situación muy personal de contactarme con el padre de mi hija porque yo soy madre soltera, que que ya tenía muchos años en España y como no tenía resuelto esa situación, eh, yo quería de una vez terminar con esta pesadilla personal y es por eso que que vine a España, no.

P: O sea, el padre de tu hija estaba aquí en España y, tú a tu hija la dejas allí...

R: Con mis padres.

P: Con tus padres.

R: Si, es la primera vez que me separo de ella pero al verme que tenía dos, retornar a Colombia, este, endeuda, y encima si un futuro claro pues decidí ...

P: Venirte para acá.

R: esperar este tener la oportunidad de mejorar acá, de progresar acá y de traer más bien a la niña acá, es lo que busco.

P: ¿Qué edad tiene la niña?

R: 13 años,

P: Una de doce, ya son mujercitas.

R: Si, ya son, ya están doraditas, ya, si, si.

P: Y ¿Te costó mucho? Me refiero económicamente, no, el ir a Suiza, tuviste que, es decir ¿El endeudamiento que eso te suponía al salir de tu país era grande?

R: Si, en total, pues, representa un valor de 5.000 francos, algo de 4. 500 euros.

P: Mucho dinero.

R: Si.

P: ¿Por qué? ¿Por lo que es el avión o por qué sube tanto?

R: Bueno, es que ese concepto es el que me lo ha impuesto mi prima, prácticamente, que ahí ella ha considerado desde papeleos, desde avión, desde estadía, desde ropa, desde este alimentación y (¿hacer tareas?) y todo y como yo, este, dado la crianza y la formación yo por cuestiones económicas no me hago problemas pues si eso es lo que tú quieres, te lo daré.

Con esta parte de la entrevista se puede realizar los pasos comentados en el apartado dedicado a análisis de datos, organizar los comentarios por temas y subtemas y luego establecer una ordenación de los mismos. Todas estas cuestiones se verán con mayor profundidad al hablar en el capítulo 10 del análisis de contenido.

BIBLIOGRAFÍA

Bosch, J. y Lluís, C. (1993). *Encuestas telefónicas y por correo*. Centro de Investigaciones Sociológicas. Madrid.

Bruner, Jerome (1991). *Actos de significado. Más allá de la revolución cognitiva.* Madrid: Alianza.

Canales, M. y Peinado, A. (1994). "Grupos de Discusión". En J. M. Delgado y J. Gutiérrez (coor.) *Métodos y Técnicas cualitativas de investigación en Ciencias Sociales*. Madrid: Sintesis.

Cohen, L. y Manion, L. (1990). *Métodos de investigación educativa.* Madrid: La Muralla.

Díaz de Rada, A. (2005). *Etnografía y Técnicas de Investigación Antropológica*. Madrid: UNED.

Flick, Uwe (2004). *Introducción a la investigación cualitativa.* Madrid: Morata.

Fontana, A. y J. H. Frey (1994). "Interviewing. The Art of Science", en N.K. Denzin y Y. S. Lincoln (eds.) *Handbook of Qualitative Research*. London: SAGE.

Goetz, J. P. y M. P. Le Compte (1986). *Etnografía y diseño cualitativo en investigación Educativa*. Madrid: Morata.

Groves, R. M. (1989). *Survey errors and survey costs*. New York, Wiley.

Hammersley, M. y Atkinson, P. (1994). *Etnografía. Métodos de investigación*. Barcelona: Paidós.

Hyman, Herbert (1984). *Diseño y análisis de las encuestas sociales*. Amorrortu. Buenos Aires.

Ibáñez, J. (1979). Más allá de la Sociología. El grupo de discusión: Técnica y Crítica. Madrid: Siglo XXI.

Igúzquiza, Vidal (2002). *Tipos de encuestas y diseños de investigación.* Universidad Pública de Navarra, D.L. Pamplona.

Kish, Leslie (1995). *Diseño estadístico para la investigación*. Siglo XXI. Madrid.

Krueger, R. A. (1988). *El grupo de discusión. Guía práctica para la investigación aplicada.* Madrid: Pirámide.

McCracken, G. (1988). *The Long Interview (Qualitative Research Methods)*. London: Sage.

Nahoum, C. (1961). *La entrevista psicológica*. Buenos Aires: Kapelusz.

Patton, H. (1980). *Qualitative Evaluation*. London: Sage.

Spradley, J. P. (1979). *The Ethnographic Interview*. New York: Holt, Rinehart and Winston.

Taylor, S. J. y Bogdan, R. (1994). *Introducción a los métodos cualitativos de investigación*. Barcelona: Paidós.

Tuckman, B. W. (1972). *Conducting Educational Research*. New York: Harcourt Brace Jovanovich.

Vázquez Recio, R. y Angulo Rasco, F. (2003). *Introducción a los estudios de casos. Los primeros contactos con la investigación etnográfica*. Archidona: Aljibe.

Walter, Rob (1989). *Métodos de investigación para el profesorado*. Madrid: Morata.

Woods, P. (1987). *La escuela por dentro. La etnografía en la investigación educativa*. Barcelona/Madrid: Paidós/MEC.

Whyte, W. F. (1982). *Learning from the field: a guide from experience*. Beverly Hills: Sage.

Capítulo 7
LAS ESCALAS

1. Las escalas
2. Escalas de actitudes
   2.1. Escalas de Likert
   2.2. Escalas de Thurstone
   2.3. Escalas de Guttman
3. Escalas de producción escolar
4. Diferencial semántico

Objetivos

- Valorar la utilidad de las escalas en las investigaciones sociales
- Conocer las distintas variantes de escalas y sus limitaciones
- Saber elegir el modelo de escala más apropiado a un contexto concreto
- Diseñar una escala
- Realizar los controles técnicos apropiados para controlar la calidad de cada escala
- Saber implementar una escala en el contexto de una investigación
- Recoger y analizar los datos recogidos de una escala

1. LAS ESCALAS

Aunque el término escala tiene un significado diverso —escala de medida, instrumento de medición o numerales de un instrumento de medida— en este punto vamos a centrarnos en el segundo concepto.

Las escalas se utilizan para determinar el grado o frecuencia con que se da una característica, rasgo o conducta. Por lo general, una escala consta de una serie de ítems y un conjunto de categorías que representan el grado de variación de la característica, rasgo o conducta que representa el ítem.

Se diferencia de la lista de control en que mientras con esta última se determina la presencia o ausencia de un rasgo, característica o conducta, a través de las escalas se intenta determinar el grado o frecuencia en que una determinada característica se presenta.

A la hora de clasificar las escalas se pueden utilizar diferentes criterios, no obstante, los más generales atienden al ***número de dimensiones de la escala*** y la ***entidad a escalar***.

Según el número de dimensiones de la escala se tiene:

a) El escalamiento unidimensional: que posiciona los diferentes objetos en una única dimensión.

b) El escalamiento multidimensional: que posiciona los objetos en dos o más dimensiones simultáneamente.

Si se atiende al segundo criterio y supuesto, una ***entidad*** es cualquier cosa que produce una respuesta en el sujeto. En función de éste criterio, propuesto por Torgerson, se puede hablar de:

a) ***Métodos centrados en los sujetos***: que estudia las diferencias individuales en las actitudes producidas en los sujetos ante determinados estímulos. Las diferentes respuestas representan una forma de escalar

a los sujetos ante la actitud medida. El método de escalamiento unidimensional, se conoce como escala sumativa o de Likert (1932), y el método multidimensional es la técnica de análisis factorial.

b) ***Métodos centrados en los estímulos***: mide y ordena los diferentes estímulos u objetos propuestos. Como técnica unidimensional estaría el método de las comparaciones binarias de Thurstone (1927, 1929) y como técnicas multidimensionales estarían el modelo métrico de Torgerson y el modelo no métrico de Shepard y Kruskal (1962, 1964). Además, otros dos modelos de escalamiento multidimensional: el Modelo INDSCAL y el Modelo con replicación, que se diferencian de los otros en que permiten partir de varias matrices de entrada frente a los clásicos que sólo permiten partir de una matriz, la matriz promedio.

c) ***Métodos centrados en las respuestas o métodos de escalamiento conjunto***: se trata de escalar los estímulos y los sujetos, es decir las respuestas. En escalamiento unidimensional estarían el escalograma de Guttman (1942) y el Modelo del punto ideal de Coombs (1950) y en escalamiento multidimensional estarían el Modelo vectorial y el Modelo del punto ideal.

Otra clasificación de las escalas es según su uso: ***las de estimación, de actitudes, de producción escolar y de diferencial semántico*** (Bisquerra, 1989, p. 107).

Cuando la escala se aplica en el contexto de la observación, ya sea ésta realizada por el investigador o sus colaboradores, se le conoce por el nombre de ***escala de estimación (rating scales)***, a través de la cual el observador aprecia el grado en que una característica o conducta se da. Entre este tipo de escalas se diferencian, a su vez, entre las numéricas, gráficas, descriptivas y, mixtas. En estas observaciones, a veces, el investigador anota observaciones pasadas que recuerda. Entre las dificultades que presenta este tipo de escala está la subjetividad en las calificaciones y la complicación de delimitar los límites entre grados. Esta subjetividad puede llevar, algunas veces, a provocar el denominado ***efecto halo***, es decir, la trasposición de una impresión a un campo diferente, por ejemplo, considerar como inteligente a un alumno que tiene buena presencia. También hay que señalar el efecto de ***tendencia central***, es decir, la predisposición del investigador a una valoración central.

Las ***escalas de actitudes (o de opinión)***, intentan medir la actitud de los sujetos a través de la valoración que cada sujeto asigna a cada ítem que compone la escala.

Tradicionalmente se suelen distinguir tres tipos de escalas de actitud: diferenciales, sumativas y acumulativas. También se conocen por el nombre de sus autores: Thurstone, Likert y Guttman, respectivamente.

En las ***escalas diferenciales (Thurstone)*** los ítems se analizan y seleccionan para cubrir todo el espectro de la variable que se desea medir, y cada uno de ellos tiene un valor distinto y, por lo tanto, las mismas respuestas a ítems diferentes tienen un peso distinto en la puntuación final. En la aplicación de la escala, los sujetos se limitan a señalar las afirmaciones con las que están de acuerdo. Entre las dificultades que presenta este tipo de escala está lo dificultoso de su construcción. Otro inconveniente que presenta es la posibilidad de que dos personas tengan la misma puntuación habiendo contestado de forma distinta.

En las ***escalas sumativas (Likert)*** los ítems no se escogen para cubrir todo el espectro, sino en función de su relación comprobada con todos los demás, es decir, su poder de discriminación, y tienden a estar en los extremos del continuo. Además, todos los ítems tienen el mismo valor y las mismas respuestas dadas a ítems distintos tienen idéntico valor. Los sujetos responden indicando su grado de acuerdo o desacuerdo con el ítem. Cada sujeto tiene como puntuación global la suma de las puntuaciones asignadas a cada ítem. Por eso se llama escala sumativa. Presenta el mismo inconveniente expuesto para la escala de Thurstone, a saber, diferentes combinaciones de puntuaciones pueden dar la misma nota final.

Finalmente, en las ***escalas acumulativas (Guttman)*** todos los ítems deben medir lo mismo aunque en grados distintos de intensidad. Son escalas con muy pocos ítems (cinco o seis), de confección laboriosa y útiles para medir objetivos muy limitados. Son escalas unidimensionales.

Las ***escalas de producción escolar*** constituyen una serie graduada de trabajos tipo, que sirven como punto de comparación para apreciar los trabajos similares realizados por cualquier sujeto (Bisquerra, 1989 p.112). Ejemplo de las mismas son las escalas de ortografía, dibujo y de composición, etc.

El ***diferencial semántico*** es un instrumento de recogida de información, a través del cual se pretende conocer los significados connotativos

que tienen los objetos, fenómenos, hechos o situaciones para los sujetos. Mediante este instrumento se pide a un sujeto que puntúe un concepto en función de varios criterios. Cada criterio se presenta como una pareja de adjetivos bipolares (bueno-malo, blanco-negro, cálido-frío, etc).

Este tipo de instrumento fue introducido a principio de los años cincuenta por Osgood, cuando estudiaba el significado de conceptos bipolares, basados en el hecho de que las personas califican conceptos utilizando adjetivos, y que éstos suelen tener sus contrarios.

A través del diferencial semántico el investigador puede: a) comprobar la diferente valoración o conceptuación que de objetos, en sentido amplio, tienen personas o/y grupos distintos; b) apreciar las semejanzas que en la conceptuación o/y la valoración de objetos se dan entre grupos o individuos semejantes o diferentes; y c) situar objetos o conceptos en categorías en función de su definición y valoración similar respecto al espacio semático (López-Barajas, 1985, p.290). Las escalas de medida obtenidas mediante el diferencial semántico, a efectos de tratamiento estadístico, se supone son de intervalo.

2. ESCALAS DE ACTITUDES

Aún cuando no es nuestro objetivo entrar en la definición de actitud, sabemos que representa la predisposición de actuar de una cierta forma o manera, en contra o a favor de algún objeto. Algunos autores encuentran tres componentes en las actitudes: *cognitivo, afectivo y conductual* o *de comportamiento*. Cuando, por ejemplo, se mide la actitud ante el consumo del alcohol, existe una componente cognitiva sobre el conocimiento, creencia, valores, ideas, etc. sobre el alcohol. También existe otra componente afectiva, es decir, los sentimientos que produce: mal, bien, indiferente... Y finalmente una componente conductual, donde se observará la conducta más probable ante el objeto; por ejemplo, lo más probable es que no beba, aunque tenga una actitud favorable al uso del alcohol. Se ha comprobado empíricamente que existe una correlación media (0,5) entre la actitud ante un objeto y la conducta producida; no obstante, como resulta más fácil la medida de actitudes que de conductas se aborda las primeras en la idea de saber las conductas más probables.

Las ***escalas de actitudes (o de opinión)***, intentan medir la actitud de los sujetos a través de la valoración que cada sujeto asigna a cada ítem que compone la escala.

2.1. Escalas de Likert

En las ***escalas sumativas (Likert)*** los ítems no se escogen para cubrir todo el espectro, sino en función de su relación comprobada con todos los demás, es decir, su poder de discriminación, y tienden a estar en los extremos del continuo. Además, todos los ítems tienen el mismo valor y las mismas respuestas, dadas a ítems distintos, tienen idéntico valor. Los sujetos responden indicando su grado de acuerdo o desacuerdo con el ítem, es decir, su ***juicio de valor*** del ítem, o de otra forma lo que ***debería ser*** no lo que sea el ítem. Cada sujeto tiene como puntuación global la suma de las puntuaciones asignadas a cada ítem. Por eso se llama escala sumativa. Presenta el inconveniente de que diferentes combinaciones de puntuaciones pueden dar la misma nota final.

El proceso de elaboración de una escala de Likert está inmerso en una serie de tareas (Barbero, 1993):

1. Preparación de los ítems iniciales.
2. Asignación de puntuaciones a los ítems.
3. Análisis de los ítems iniciales para su selección.
4. Evaluación de la escala definitiva.
5. Fiabilidad y validez de la escala

La escala Likert es, en estricto sentido, una medición ordinal, sin embargo, se suele trabajar como si fuera de intervalo.

Preparación de los ítems iniciales

En primer lugar se debe estudiar la dimensión de la actitud a medir y las conductas relacionadas con la misma. De este proceso se recogerán un conjunto de ítems, mayor que los que compondrán la escala final. El número de ítems elegido permitirá cubrir todo el rango de conductas que refleja la actitud en estudio.

Se cuidará en la redacción del ítem que la persona pueda emitir su juicio de valor a la pregunta formulada. Así un ejemplo de ítem puede ser: ¿Considera usted que deberían los alumnos leer en casa para fomentar la compresión lectora?

( ) Indispensable

( ) Muy importante

( ) Medianamente importante

( ) Poco importante

( ) No es importante

El orden de presentación de las preguntas debe ser aleatorio para no producir sesgos por la presencia de rachas de preguntas que reflejen actitudes positivas o favorables y otras sucesiones de preguntas de actitudes negativas.

Se administrará los ítems iniciales a una muestra piloto de individuos, para desechar ítems mal formulados o de baja discriminación en la escala.

Asignación de puntuaciones a los elementos y a los sujetos

Existe dos cuestiones a debatir: número de categorías que tendrán los ítems y asignación de numerales a las distintas categorías del ítem.

Respecto a la primera cuestión, Likert en el estudio que le llevó a la creación de la escala de su nombre aconsejó que los ítems de la escala tuvieran 5 categorías. Las más utilizadas son:

a) Completamente de acuerdo

b) De acuerdo

c) Indiferente

d) En desacuerdo

e) Completamente en desacuerdo

Otros autores desaconsejan ésta escala y añaden más categorías. Por la facilidad de aplicar pruebas estadísticas, el número de categorías aconsejable será aquel cuya distribución de frecuencias produzca mayor proximidad a la distribución normal. Otra cuestión a tener en cuenta será la facilidad de deci-

sión entre las distintas opciones o categorías del ítem; así se desaconseja ítems con muchas categorías.

La segunda cuestión es la asignación de numerales a cada una de las categorías del ítem. Se suelen emplear dos estrategias:

- Asignar el 1 a la categoría completamente en desacuerdo y el 5 a la completamente de acuerdo e ir incrementando una unidad. Así quedaría la asignación de la siguiente forma:

  a) Completamente de acuerdo 5
  b) De acuerdo 4
  c) Indiferente 3
  d) En desacuerdo 2
  e) Completamente en desacuerdo 1

- Asignar el 0 a la categoría central e ir incrementando o decrementando una unidad según se suba o se baja el nivel de acuerdo con el ítem. Así quedaría:

  a) Completamente de acuerdo 2
  b) De acuerdo 1
  c) Indiferente 0
  d) En desacuerdo ..... –1
  e) Completamente en desacuerdo ..... –2

Habrá que tener presente en la asignación de numerales a las categorías las preguntas formuladas en sentido negativo, donde se deberá invertir la asignación; entonces completamente de acuerdo pasará a valor 1 ó –2 ,según los casos, y completamente en desacuerdo 5 ó 2 respectivamente.

Análisis de los ítems iniciales para su selección

Para estudiar el poder discriminatorio de cada ítem en la escala Likert propuso dos procedimientos:

a) Estudiar el grado de homogeneidad del ítem medido por el coeficiente de correlación entre las puntuaciones de los sujetos en el elemento y el resto de los elementos (el total de la escala, menos el ele-

mento en estudio). Coeficientes de correlación bajos, inferiores a 0,2 serán indicadores de falta de homogeneidad con la escala y deberán eliminarse de la misma.

b) Controlar la discriminación de cada elemento mediante el estudio de las diferencias significativas entre los sujetos que puntúan más alto en la escala (habitualmente se elige el 25%) y los que puntúan más bajo (el otro 25%, extremo). Se estudia estadísticamente si la diferencia entre las medias de los individuos de cada grupo, para el ítem en estudio, es significativa. Si no lo fuera, será indicador de que el ítem no discrimina, y en consecuencia se debe eliminar. El análisis de estas diferencias se hará con la prueba t de Student o la prueba U- de Mann-Whitney, cuando no se cumplan las condiciones paramétricas con los grupos de individuos. También se puede plantear una prueba de ji-cuadrado en una tabla de contingencia como la siguiente, una vez ordenadas las puntuaciones de los sujetos y obtenida su mediana:

	+ Mediana	**– Mediana**
Grupo superior (25%)	a	b
Grupo inferior (25%)	c	d

Likert utilizó esta última prueba en la construcción de su escala.

Evaluación de la escala. Interpretación de puntuaciones

Como se ha comentado la escala de Likert es una escala sumativa, en consecuencia para obtener la puntuación de un individuo se suma las puntuaciones en todos los ítems de la escala. No obstante como el objetivo de la escala Likert es observar las diferencias individuales, entonces interesa conocer cómo se comporta el individuo en relación con su grupo. Para éste fin se aconseja obtener las puntuación tipificadas o normalizadas (puntuaciones z), para entonces poder calcular el porcentaje de personas cuya puntuación es inferior a un determinado z. Así por ejemplo, un individuo con una z=1 tendrá aproximadamente el 84% de los casos inferiores a ese valor (véase la tabla de la distribución normal).

Como hemos podido observar para tener una idea relativa de las diferencias individuales, las puntuaciones de un sujeto siempre tienen que refe-

renciarse en su grupo. En la actualidad el desarrollo de los modelos TRI (Teoría de respuesta al ítem) no necesitan la referencia al grupo para calcular las puntuaciones de los sujetos y las diferencias entre los mismos.

Fiabilidad y validez de la escala

Los métodos más utilizados para el cálculo de fiabilidad de la escala de Likert son: el método de las dos mitades con la correlación de Spearman-Brown, y la alfa de Cronbach.

Existen distintos indicadores de validez utilizables en esta escala: de constructo, predictiva, etc. Remitimos al lector a los procedimientos comentados en el capítulo 2.

2.2. Escalas de Thurstone

En las ***escalas diferenciales (Thurstone)*** los ítems se analizan y seleccionan para cubrir todo el espectro de la variable que se desea medir, y cada uno de ellos tiene un valor distinto y, por lo tanto, las mismas respuestas a ítems diferentes tienen un peso distinto en la puntuación final. En la aplicación de la escala, los sujetos se limitan a señalar las afirmaciones con las que están de acuerdo. Entre las dificultades que presenta este tipo de escala está lo dificultoso de su construcción. Otro inconveniente que presenta es la posibilidad de que dos personas tengan la misma puntuación habiendo contestado de forma distinta.

El proceso de construcción de la escala de Thurstone comprende:

- Definir la variable de actitud que se quiere medir
- Preparar los ítems iniciales
- Clasificar los elementos en categorías: Prueba de Jueces
- Criterios para la selección de los elementos:
  - — Valor escalar de los elementos
  - — Ambigüedad e irrelevancia
- Obtención de la escala definitiva
- Fiabilidad y validez de la escala

Definición de la variable

La definición de la variable actitud que se quiere medir comprende una revisión bibliográfica de la literatura existente sobre la actitud objeto. Esta revisión permitirá definir un conjunto de características, componentes del continuo de la escala unidimensional que define la actitud, distintos grados de la actitud, que posibilitará generar los ítems.

Preparación de los ítems iniciales

Para preparar los ítems iniciales, se recogerán, sin restricciones, cuantos ítems se crean necesarios para cubrir todas las características de la actitud objeto. Se cogerán ítems que representen posiciones positivas, negativas e indiferentes de la actitud. Se suelen proponer unos 100 ítems iniciales.

Clasificación de los elementos en categorías: prueba de Jueces

La clasificación de los elementos en categorías supone la ordenación de los enunciados según el grado de actitud que denotan. Esta graduación se suele hacer en una escala de 11 niveles y su valoración la realiza una muestra representativa de la población, compuesta por unos 200 jueces que gradúan el nivel que denota cada ítem como componentes de la actitud.

Al clasificar los jueces el continuo de actitud en 11 intervalos iguales de amplitud 1. El intervalo neutro será el (5,5-6,5), el que indica una actitud más desfavorable sería el (0,5-1,5) y el que indica una actitud más favorable el (10,5-11,5).

Criterios para la selección de los elementos

De los aproximadamente 100 ítems iniciales se debe pasar a unos 25-40 elementos, por tanto debe existir un proceso de selección con unos criterios objetivos de filtrado. Thurstone propone los siguientes:

a) Valor escalar de los elementos

La idea que se persigue es conseguir un conjunto de elementos formando una escala graduada de la actitud estudiada. Por tanto,

deberán estar presentes, como mínimo, 2 o 3 ítems por cada uno de los intervalos del continuo de la actitud.

El valor de un ítem en la escala será igual a la mediana de las valoraciones de los jueces. Se deberá conseguir que los valores escalares de los distintos ítems de la escala cubran por completo todos los valores de la misma.

b) Criterio de ambigüedad

Este término está relacionado con el grado de consenso logrado por los distintos jueces a la hora de evaluar un ítem.

Un indicador de consenso es el llamado coeficiente de ambigüedad=C.A.=Q_3-Q_1=diferencia entre el tercer y el primer cuartil, para un determinado ítem. Si éste valor es superior a 2 será indicador de falta de coincidencia entre jueves y el ítem debe ser eliminado. Cuando los jueces puntúan en el rango promedio (5,5-6,5) esta diferencia debe ser mayor que 3.

Así por ejemplo, sean los resultados de un elemento E en la valoración de la actitud realizada por 300 jueces.

Categorías E	n_i	N_i
1	0	0
2	0	0
3	0	0
4	50	50
5	180	230
6	40	270
7	30	300
8	0	300
9	0	300
10	0	300
11	0	300

$$Q_1 = 4,5 + \frac{75-50}{180} = 4,64 \quad Q_3 = 4,5 + \frac{225-50}{180} = 5,47$$

$$\text{CA} = 5,47 - 4,64 = 0,83$$

c) Criterio de irrelevancia

Se puede dar el caso que algunos ítems incluyan connotaciones que quedan fuera del rango de aspectos de la actitud medida. Estos ítems deben ser eliminados.

Para la selección de los ítems mencionados, se cogerá un grupo de jueces o sujetos y se les pedirá si apoyan (+) o rechazan ( ) cada uno de los ítems porque no coinciden con su actitud. Entonces para dos ítems a y b se puede calcular un indicador de la semejanza entre los mismos como:

$$\text{I.S.(ab)} = \frac{N_{ab}}{N_b} \text{ donde}$$

N_{ab}=núm.de sujetos que responden positivamente a los ítems a y b

N_b=núm.de sujetos que responden positivamente al ítem b

El índice oscila entre 1, máxima semejanza, y 0, mínima similitud.

Este proceso de comparación se debe realizar cada ítem con todos los demás y sus valores irán decreciendo conforme más alejados estén en valoración. Si existe un ítem que no goza de éstas característica y sus valores de I.S. permanecen constantes en sus comparaciones con el resto de ítems, debe ser eliminado.

Obtención de la escala definitiva

La escala definitiva estará conformada por el conjunto de ítems que han superado las pruebas anteriores. Para su presentación en el cuestionario se aleatorizará el orden. Cada uno de los ítems tendrá dos alternativas para que los sujetos manifiesten su acuerdo o descuerdo con el contenido del mismo.

La escala es de aplicación individual o colectiva. La puntuación de cada sujeto vendrá dada por la media de los valores escalares asociados a los elementos seleccionados favorablemente.

Fiabilidad y validez de la escala

Los métodos más utilizados para el cálculo de fiabilidad de la escala de Thurstone son: el método de las dos mitades con la correlación de Spearman-Brown, y la alfa de Cronbach.

Para que la escala sea válida los valores escalares asignados por los sujetos a los elementos deberá ser independiente de la muestra de sujetos elegida. Una forma de obtener la validez de la escala será aplicar la misma a dos muestras independientes, a ser posible antagónica, y comprobar si no hay diferencia significativa en los valores asignados a los distintos ítems de la escala.

2.3. Escalas de Guttman

En las ***escalas acumulativas (Guttman)*** todos los ítems deben medir lo mismo aunque en grados distintos de intensidad. Son escalas con muy pocos ítems (cinco o seis), de confección laboriosa y útiles para medir objetivos muy limitados. Son escalas unidimensionales.

El Escalograma de Guttman se basa en la idea de que si un sujeto responde favorablemente a un determinado estimulo lo hará también a todos aquellos que estén por debajo de él en la escala de inclusión construida.

Pensemos, por ejemplo, que se han escogido un conjunto de preguntas de Estadística Aplicada y se han ordenado según su grado de dificultad. Si estas preguntas ordenadas forman una escala de Guttman, entonces un sujeto que responda a una determinada pregunta, acertará todas las preguntas más sencillas y fallará todas las preguntas de mayor dificultad.

En consecuencia se pueden ordenar tanto los sujetos como los estímulos (preguntas de Estadística Aplicada), en un continuo que forma el escalograma de Guttman.

En esencia la escala de Guttman además de un procedimiento de medida de actitudes es una técnica para comprobar si un conjunto de ítems forman una escala unidimensional, indicadores de un constructo subyacente.

Se supone un conjunto de 6 sujetos que han respondido a 5 ítems o elementos dicotómicos, indicando 1 si acierta el ítem y 0 si lo falla. Se forma la siguiente tabla:

Sujetos	Ítems					P. Escala
	1	2	3	4	5	
S_1	1	1	1	1	1	5
S_2	1	1	1	1	0	4
S_3	1	1	1	0	0	3
S_4	1	1	0	0	0	2
S_5	1	0	0	0	0	1
S_6	0	0	0	0	0	0
P. Item	5	4	3	2	1	

La tabla es un ejemplo de una escala de Guttman perfecta donde sabiendo la posición de un sujeto se puede saber exactamente el numero de respuesta que ha acertado. Por ejemplo el sujeto S_4 ha respondido a 2 preguntas y estas sólo pueden ser las preguntas 1 y 2. Luego se puede comprobar que éste procedimiento ordena los sujetos y los ítems en un continuo.

Como se ha podido intuir por el ejemplo resulta difícil construir una escala de Guttman perfecta, por tanto el problema será qué grado de desviación de la escala perfecta se admite y que indicadores se utilizarán para medir tal desviación.

La capacidad de poder indicar a partir de una puntuación de un sujeto qué ítems acepto y qué ítems rechazo, se conoce como reproductividad. Para controlar las desviaciones que se producen por errores de los sujetos o por la falta de unidimensionalidad de los ítems, se utiliza el coeficiente de reproductividad:

$$\text{C.R.} = 1 - \frac{\text{núm. total de errores}}{\text{núm. total de respuestas}} = 1 - \frac{\text{núm. total de errores}}{\text{núm. total de sujetos} \times \text{núm. ítems}}$$

Por convenio se considera que un conjunto de ítems constituyen una escala de Guttman si su coeficiente de reproductividad es mayor que 0,90.

Para aclarar el cálculo del coeficiente C.R. veamos un ejemplo:

Sujetos	Ítems					P. Escala
	1	2	3	4	5	
S_1	1	1	1	1	0	4
S_2	1	1	0	1	0	3
S_3	1	1	1	0	1	4
S_4	1	0	0	1	0	2
S_5	1	0	0	0	1	2
S_6	0	1	0	1	0	2
S_7	1	1	1	1	1	5
S_8	1	0	1	0	0	2
S_9	0	0	0	0	1	1
P. Item	7	5	4	5	4	

Se ordenan los sujetos y los ítems, con la opción que presente menos errores, dando como resultado la tabla:

Sujetos	Ítems					P. Escala	Núm. errores
	1	2	3	4	5		
S_1	1	1	1	1	1	5	0
S_2	1	1	1	1	0	4	0
S_3	1	1	1	1	1	4	2
S_4	1	0	1	0	0	3	0
S_5	1	1	0	0	0	2	0
S_6	1	0	0	0	1	2	2
S_7	0	1	1	0	0	2	2
S_8	1	0	0	1	0	2	2
S_9	0	0	0	0	1	1	2
P. Item	7	5	5	4	4		10

Para comprender el cálculo de los errores supongamos un sujeto, por ejemplo el S_4 , su respuesta es (1 0 0 1 0), sin embargo el patrón de respuesta ideal es (1 1 0 0 0), por tanto el número de transformaciones que habría que hacer en su respuesta para ajustarse al patrón ideal serían 2. Finalmente el C.R. será:

$$\text{C.R.} = 1 - \frac{10}{45} = 0,77$$

Si la escala es perfecta o el C.R. es mayor o igual a 0,90, la puntuación de un sujeto se obtiene sumando el número de respuestas favorables a los ítems. Cuando no se da esta situación, como en el ejemplo propuesto donde el CR es menor de 0,90, deberá modificarse la escala para ajustarse a la ideal.

3. ESCALAS DE PRODUCCIÓN ESCOLAR

Las ***escalas de producción escolar*** constituyen una serie graduada de trabajos tipo, que sirven como punto de comparación para apreciar los trabajos similares realizados por cualquier sujeto (Bisquerra, 1989 p.112). Ejemplo de las mismas son las escalas de ortografía, dibujo y de composición, etc.

4. DIFERENCIAL SEMÁNTICO

El ***diferencial semántico*** es un instrumento de recogida de información, a través del cual se pretende conocer los significados connotativos, afectivos o subjetivos, que tienen los objetos, fenómenos, hechos o situaciones para los sujetos. Mediante este instrumento se pide a un sujeto que evalúe un concepto en función de varios criterios. Cada criterio se presenta como una pareja de adjetivos bipolares (bueno-malo, blanco-negro, cálido-frío, etc).

Este tipo de instrumento fue introducido a principio de los años cincuenta por Osgood, cuando estudiaba el significado de conceptos bipolares, basados en el hecho de que las personas califican conceptos utilizando adjetivos, y que éstos suelen tener sus contrarios.

A través del diferencial semántico el investigador puede: a) comprobar la diferente valoración o conceptuación que de objetos, en sentido amplio, tienen personas o/y grupos distintos; b) apreciar las semejanzas que en la conceptuación o/y la valoración de objetos se dan entre grupos o individuos semejantes o diferentes; y c) situar objetos o conceptos en categorías en función de su definición y valoración similar respecto al espacio semántico (López-Barajas, 1985, p. 290). Las escalas de medida obtenidas mediante el diferencial semántico, a efectos de tratamiento estadístico, se supone son de intervalo.

Para construir el diferencial semántico será necesario realizar el siguiente proceso:

1. Definir el concepto a medir
2. Definir las escalas bipolares
3. Aplicación del diferencial semántico
4. Análisis de datos

Definir el concepto

Es el objeto a evaluar por el sujeto. Deberá ser definido con precisión para evitar ambigüedades.

El estimulo u objeto puede ser muy variado incluyendo conceptos verbales (palabras) o no verbales (imágenes, estímulos físicos).

Del conjunto de conceptos que puede definir la variable actitudinal en estudio, será preciso realizar el correspondiente filtrado con criterios de representatividad y relevancia, de tal manera que permita: a) la discriminación entre sujetos, b) claridad y unicidad para el individuo, y c) familiaridad para el mismo.

Definir las escalas bipolares

a) Escalas bipolares

Cada uno de los conceptos elegidos para definir la variable se evalúa mediante las escalas bipolares. Éstas representan una reacción de tipo afectivo hacia el concepto: bueno-malo, activo-pasivo, agradable-desagradable, lento-rápido, etc.

Se representan en un continuo, dividido normalmente en siete categorías, cuyos extremos son los adjetivos mencionados:

1. malo	1	2	3	4	5	6	7 bueno
2. pasivo	1	2	3	4	5	6	7 activo
3. desagradable	1	2	3	4	5	6	7 agradable
4. lento	1	2	3	4	5	6	7 rápido

.....

b) Selección de las escalas

El número de escalas bipolares que se puede utilizar para definir un concepto es ilimitado. Sin embargo, por operatividad, es necesario acotar dicho número, para lo cual en la construcción de las escalas bipolares se coge una muestra de sujetos y se les pide detallen cuantos adjetivos se relacionen con el concepto presentado; luego se analizan dichos adjetivos y se seleccionan aquellos que han aparecido con mayor frecuencia. Por ejemplo, en los trabajos de Osgood la escala definitiva estuvo constituida por 20 conceptos con 50 escalas bipolares cada uno.

Una técnica estadística apropiada para encontrar los factores subyacentes al significado semántico de los conceptos a evaluar es el análisis factorial. En la selección de las escalas se procurará haya tres o cuatro escalas bipolares para cada factor, con saturaciones altas en el factor representado y bajas o nulas para el resto de los factores.

Otro criterio de selección de las escalas será la **relevancia** de las mismas. Porque puede haber escalas con altas saturaciones pero poco o nada relevantes para el concepto a evaluar.

Finalmente, otro criterio de selección es la **estabilidad semántica** de la escala en relación con los conceptos y sujetos de la investigación. Por ejemplo la escala grande-pequeño puede ser apropiada para evaluar conceptos como: león, barco, casa, etc., mientras será inapropiada para evaluar conceptos como: Patria, Ministro, Cura, etc.

Aplicación del diferencial semántico

En la elaboración del diferencial semántico se seguirá el siguiente proceso:

1. Elaborar lista de los conceptos a evaluar
2. Selección previa de los adjetivos a utilizar para evaluar la lista según el criterio de máxima frecuencia
3. Buscar las palabras que tengan el significado opuesto a los adjetivos encontrados
4. Selección de las escalas definitivas según los criterios comentados anteriormente

Una vez elaborado el diferencial semántico su aplicación se suele hacer en cuadernillos, donde aparecen los distintos conceptos con sus escalas bipolares correspondientes; se suele utilizar una hoja para cada concepto y se pide a los sujetos que contesten de forma rápida, ya que prima la espontaneidad.

Para la valoración de cada escala bipolar se tendrá en cuenta que el número más pequeño de la escala deberá corresponder a la categoría más próxima al adjetivo que represente una valoración negativa del concepto y el número más alto, corresponderá a la categoría más próxima al adjetivo que refleja una valoración positiva.

La puntuación de cada sujeto en cada escala será el valor numérico asignado a la categoría elegida dentro del continuo de 1 a 7 o de –3 a 3 (según la elección del investigador).

Análisis de datos

Con las respuestas obtenidas se realiza los siguientes análisis estadísticos:

1. Con las puntuaciones escalares:
   - La media de la muestra en cada una de las escalas
   - La media de todas las escalas que evalúan un concepto determinado, tanto para el individuo como para el grupo
2. Puntuaciones factoriales

   Al realizar el análisis factorial de las puntuaciones escalares se obtienen las puntuaciones factoriales. Los trabajos de Osgood pusieron de manifiesto la existencia de tres dimensiones fundamentales: *evaluativa, potencia y actividad*. La primera de mayor importancia en la explicación de la variación de las puntuaciones.
3. Análisis de perfiles

   Osgood y Luria pensaron que la distancia entre las puntuaciones factoriales podía dar una idea de las diferencias entre grupos, sujetos o entre el mismo sujeto en momentos distintos. Esta distancia según Osgood será la distancia euclídea entre las puntuaciones factoriales:

$$\mathrm{D(ab)} = \sqrt{\Sigma d^2(a,b)}$$

Veamos un ejemplo de su cálculo:

Supongamos los siguientes datos de puntuaciones factoriales de cuatro conceptos: Salud, Felicidad, Amor, Estado en los tres factores de Osgood:

Conceptos	Evaluativa	Potencia	Actividad
Salud	5	3	6
Felicidad	3	2	2
Amor	4	5	4
Estado	4	4	5

Se puede construir la matriz de distancias euclídeas:

Conceptos	Salud	Felicidad	Amor	Estado
Salud	—			
Felicidad	4,58	—		
Amor	3,00	3,74	—	
Estado	1,73	3,74	1,41	—

Donde por ejemplo:

$$D\ (\text{salud}, \text{felicidad}) = \sqrt{(5-3)^2 + (3-2)^2 + (6-2)^2} = 4{,}58$$

Con estas distancias se puede hacer una representación gráfica, donde comenzando por cualquier punto, por ejemplo salud, se puede representar en el plano el resto de los puntos con la condición de respetar la distancia entre los mismos. También utilizando los valores de las distancias se puede hacer inferencias sobre la significatividad o no de tales diferencias.

BIBLIOGRAFÍA

BARBERO, M.I. (1993). *Psicometría II. Métodos de elaboración de escalas.* Madrid: UNED.

BISQUERRA, R. (1989). *Métodos de Investigación educativa. Guía práctica*. Barcelona: Ceac.

BRIONES, G. (1982). *Métodos y técnicas de investigación para las ciencias sociales.* Mexico: Trillas.

BUNGE, M. (1976). *La investigación científica*. Barcelona: Ariel.

DEL RINCÓN, D. y otros (1995). *Técnicas de investigación en Ciencias Sociales*. Madrid: Dykinson.

DUVERGER, M. (1974). *Método de las ciencias sociales*. Barcelona: Ariel.

FONTANA, A. y FREY, J. H. (1994). "Interviewing: The Art of Science". En N, K. Denzin y Y. S. Lincoln (Eds.), *Handbook of Qualitative Research* (pp. 361-376). Londres: Sage.

GARCÍA CUETO, E. (1993). *Introducción a la Psicometría*. Madrid: Siglo XXI de España.

GÓMEZ BENITO, J. (1986). *Los Modelos Causales como Metodología de Validez de constructo*. Barcelona: Alamex.

KERLINGER, F. N. (1985). *Investigación del comportamiento. Técnicas y metodología*. México: Interamericana.

LÓPEZ FEAL, R. (1986). *Construcción de instrumentos de medida en Ciencias Conductuales y Sociales.* Barcelona: Alamex.

MARTÍNEZ ARIAS, M. R. (1995). *Psicometría: Teoría de los Tests Psicológicos y Educativos*. Madrid: Síntesis.

MORALES VALLEJO, P. (1988). *Medición de actitudes en Psicología y Educación*. San Sebastián: Tártalo.

RINCÓN, D. del; ARNAL, J.; LATORRE, A. y SANS, A. (1995). *Técnicas de investigación en Ciencias Sociales.* Madrid: Dykinson.

SIERRA BRAVO, R. (1985). *Técnicas de investigación social. Teoría y ejercicios*. Madrid: Paraninfo.

Silva, F. (1989). *Evaluación Conductual y Criterios Psicométricos*. Madrid: Pirámide.

Visauta, S. (1989). *Técnicas de investigación social*. Barcelona: PPU.

Capítulo 8
PRUEBAS CRITERIALES

1. Pruebas criteriales
2. Tipos de pruebas criteriales
3. Elaboración de pruebas criteriales
   3.1. Especificación del dominio educativo
   3.2. Análisis de los ítems
   3.3. Determinación de estándares y puntos de corte
   3.4. Estudio de las características técnicas de la prueba
4. Un ejemplo de prueba criterial

Objetivos

- Valorar la utilidad de las pruebas criteriales en las investigaciones sociales
- Conocer las distintas variantes de las pruebas criteriales y sus limitaciones
- Saber elegir el modelo de prueba criterial más apropiado a un contexto conceto
- Diseñar una prueba criterial
- Realizar los controles técnicos apropiados para controlar la calidad de las pruebas criteriales
- Saber implementar una prueba criterial en el contexto de una investigación
- Recoger y analizar los datos recogidos de una prueba criterial

1. PRUEBAS CRITERIALES

Un grupo de test que merecen una medición aparte por sus implicaciones educativas es el de los ***tests referidos al criterio*** o test de referencia criterial.

Los tests clásicos también se llaman tests referidos a una norma, porque intentan escalar a las personas en función del grupo normativo. Los tests referidos a un criterio (TRC) van dirigidos a evaluar un criterio o ***dominio de interés***, se pone énfasis en evaluar en qué medida cada sujeto domina el criterio de interés; pero, indudablemente, los TRC también ordenan a los sujetos mediante un simple recuento del número de cuestiones respondidas bien.

Los TRC tienen sus orígenes en un artículo de Robert Glaser en 1963 en la revista American Psychologist (Muñiz, 2000, p. 99). Berk (1998, p. 365) define los tests de referencia al criterio como "construidos para evaluar el nivel de ejecución de un sujeto con respecto a un dominio de conductas bien definidas". A su vez Hambleton (Muñiz, 2000, p. 100) menciona seis campos impulsados por los TRC:

a) Un primer efecto muy positivo ha sido el obligar a profesores y constructores de tests a definir con mayor claridad y operatividad los objetivos o criterios de interés, en la línea de la evaluación conductual, para así poder construir los tests correspondientes para su evaluación.

b) Obligan a muestrear exhaustivamente los objetivos a evaluar (validez de contenido) y exigen sumo cuidado a la hora de confeccionar los ítems. En consecuencia aparece un variado abanico de formatos alternativos al omnipresente de elección múltiple.

c) Se potenciaron nuevas formas para evaluar la fiabilidad y validez de los tests (coeficiente kappa), pues las utilizadas para los tests referidos a normas no siempre resultaban las más apropiadas.

d) Dado que con gran frecuencia el uso de los TRC exigía dividir a los sujetos en dos grupos, los que dominaban el criterio y los que no, se desarrolló toda una tecnología de medida para establecer de un modo adecuado los ***puntos de corte*** para determinar quién pasa y quién falla. Indudablemente el establecer un punto de corte o criterio debe realizarse por jueces expertos en el tema objeto, adecuados en número y habituados al diagnóstico. El punto de corte debe establecerse en función de la materia y no de las personas evaluadas.

e) Los TRC, al centrarse operativamente en los objetivos específicos, han sido altamente beneficiosos para el diagnóstico de las deficiencias de aprendizaje. Permiten detectar los puntos fuertes y débiles de los sujetos y ayudan a los profesores a tomar decisiones sobre la enseñanza. Además, fomentan que los profesores hagan más hincapié en el dominio que los estudiantes tienen de la materia que en el mero análisis de las diferencias entre ellos.

f) Finalmente, ha hecho que los profesores adquieran conocimientos en el campo de la evaluación de los estudiantes. Esto es de suma importancia, pues con demasiada frecuencia algo tan relevante como la evaluación adecuada de los estudiantes se deja al sentido común de los profesores. Y el profesor, como cualquier persona, se puede confundir y perjudicar a terceros, los estudiantes, que tienen el derecho de ser evaluados con justicia para conseguir la enseñanza de calidad que todos deseamos.

A la hora de utilizar los TRC conviene recordar los peligros derivados de una elaboración inadecuada:

- Se debe cuidar la objetividad de las puntuaciones
- Se debe garantizar la validez de contenido de la prueba

2. TIPOS DE PRUEBAS CRITERIALES

Las pruebas criteriales se clasifican en función del ***dominio instruccional*** de referencia en:

- **Test de certificación o admisión**. Cuyo fin es decidir sobre la consecución o no del dominio instruccional de un nivel educativo.

- **Test de nivel o de dominio**. Para decidir sobre la adquisición o no del nivel instructivo necesario para promocionar de un curso a otro.
- **Test de aula o clase**. Su ámbito de aplicación se circunscribe al aula escolar.
- **Test de diagnóstico**. Con un ámbito más restrictivo que el anterior, pretende detectar la existencia de problemas relacionados con el aprendizaje, aportando información sobre su origen y características.
- **Test individualizado**. Representa el mayor nivel restrictivo al centrarse en un sujeto concreto, recogiendo información tanto de aspectos procesuales como de rendimiento del mismo.

3. ELABORACIÓN DE PRUEBAS CRITERIALES DE CLASE

De los diferentes tipos de pruebas criteriales nos vamos a centrar en las pruebas criteriales de clase por ser las de mayor aplicación por el profesor. La secuencia de la elaboración de las mismas puede ser la siguiente:

3.1. Especificación del dominio educativo

El conocimiento del dominio instruccional es crucial porque está relacionado con la validez de contenido y constructo (J.M. Jornet y J.M. Suarez, 1989, p. 237). Siguiendo a estos autores por dominio de instrucción se entiende el conjunto de todos aquellos elementos (objetivos, acciones, tareas, ítems) que representan el propósito de la instrucción. Las unidades que definen el dominio no deben solaparse y deben cubrir el dominio en su totalidad. Por tanto, el profesor ante una unidad didáctica o una lección debe proponer ítems que recojan todos los objetivos, contenidos y actividades relativos a la unidad a evaluar. Jornet y Suárez (1989, p. 243) recogen de Roid una clasificación de las diferentes tecnologías de generación de ítems:

Tabla 8.1. Clasificación de los procedimientos y métodos de formulación y escritura de ítems (Jornet y Suárez 1989, p. 243, tomado de Roid)

Procedimientos	Formatos de ítems Transformación lingüística
Métodos basados en teorías	Diseño de facetas Conceptos, reglas y principios Aproximación de Markle y Tiemann Sistema LOGIQ Inventario de calidad Instruccional (IQI)
Procedimientos orientados al diagnóstico	Algoritmo de Scandura Métodos basados en estrategias cognitivas

Formatos de ítems: son aplicables cuando se puede realizar un análisis de tareas o donde se puede efectuar una definición jerarquizada del dominio de instrucción. Facilita la confección de programas de ordenador "ad hoc" para ayuda en la composición e impresión de la prueba.

Transformaciones lingüísticas: permite generar ítems de acuerdo a una secuencia de cuatro puntos:

a) Búsqueda de frases relevantes en el proceso de instrucción

b) Selección de las frases más destacadas

c) Transformación de la frase según el tipo ítem, de acuerdo a la taxonomía de Bloom: reconocimiento, comprensión, aplicación y análisis

d) Construcción de distractores en los formatos de elección múltiple

Diseño de facetas: transformación de los objetivos operativos en sentencias directrices que constan de dos tipos de elementos:

a) Formulación fija de una parte del ítem

b) Formulación de una parte variable

c) La combinación de la parte fija con la parte variable da lugar a diferentes tipos de ítems.

Conceptos, reglas y principios: de los tres modelos que figuran en la clasificación de Roid, el de Tiemann y Markle supone una generación de elementos basado en la utilización conjunta de ejemplos y contraejemplos de

un determinado concepto. Los dos últimos se basan en la metodología IQI que posibilita la categorización de objetivos e ítems permitiendo la producción de éstos últimos.

Procedimientos orientados al diagnóstico: la algorítmica de Scandura, que propone analizar, mediante diagrama de flujo, los procesos implicados en la resolución de un problema. Los métodos basados en estrategias cognitivas permiten analizar los procedimientos utilizados en la resolución de problemas o la forma de actuar de alumnos exitosos y no exitosos. El acercamiento cognitivo permitirá la construcción de ítems donde se podrá interpretar las ejecuciones de los sujetos.

3.2. Análisis de ítems

Los ítems generados se pueden analizar desde dos perspectivas:

Revisión lógica: se estudia la correspondencia entre los ítems y la operativización del dominio de contenidos. Se realiza mediante el juicio de expertos quienes analizan:

a) Congruencia ítem-objetivo: donde se estudia la representatividad del ítem respecto del objetivo evaluado

b) Calidad técnica: se analiza el cumplimiento de las especificaciones técnicas sobre la medida

c) Sesgo: estudia la neutralidad de los ítems respecto a los grupos donde se pueden aplicar. Evita cualquier tendencia que favorezca o perjudique a los mismos

Revisión empírica: se estudia la aplicación de la prueba y su funcionamiento. Comprende las siguientes etapas:

a) Formulación de las hipótesis en relación con cada ítem, expresado en términos del grado de dificultad previsto y el poder discriminativo del mismo sobre el nivel de competencia establecido

b) Aplicación piloto a una muestra de alumnos

c) Análisis de los resultados de la aplicación para determinar el nivel de dificultad, discriminación y validez de cada ítem

d) Interpretación de los resultados para observar si se cumplen las hipótesis planteadas para cada ítem. Si no se obtienen los resultados esperados será necesario estudiar la especificación del dominio y la formulación de los ítems generados

En relación con la revisión empírica de los ítems es necesario matizar algunas diferencias respecto del estudio de los mismos en los test referidos a una norma (TRN).

1) El índice de dificultad (p_i) será la proporción de personas que contestan correctamente el ítem.

2) Índices de discriminación. En este caso existen diversas cuestiones diferenciadoras con los ítems de los TRN. Se pretende que los ítems de los TRC deben maximizar la discriminación de los individuos de los grupos experimental y control y minimizar la discriminación dentro del propio grupo. Estos índices también se utilizan en evaluación de programas y se conocen como **índices de sensibilidad a la instrucción**. Por resumir, se presenta dos de los índices más utilizados:

DIS_{GN} o ganancia neta. Proporción de sujetos que responden al ítem incorrectamente en el pretest y correctamente en el postest, menos la proporción de sujetos que responden incorrectamente en el pretest y en el postest. Su rango de variación es [–1,1].

$DIS_{CI\text{-}SI}$ o diferencia entre los grupos con instrucción y sin instrucción. Diferencia entre la proporción de sujetos que aciertan el ítem en el grupo con instrucción menos la proporción de individuos que aciertan el ítem en el grupo sin instrucción.

Para aclarar conceptos veamos un ejemplo con un par de ítems:

Tabla 8.2. Resultados de un par de ítems de un ejemplo simulado

	Pretest (caso 1) **Grupo Sin Instrucción (caso 2)**		**Postest (caso 1)** **Grupo Con Instrucción (caso 2)**	
Sujetos (caso 1) **Sujetos emparejados (caso 2)**	**Ítem**		**Ítem**	
	1	**2**	**1**	**2**
1	0	1	0	1
2	0	0	1	1
3	1	0	1	1
4	0	0	1	1
5	1	0	1	1
6	0	0	0	0
7	0	1	1	1
8	1	0	1	0
9	0	0	0	1
9	0	0	0	1

Los resultados de los dos ítems para los índices de dificultad y discriminación serán:

Tabla 8.3. Índices de dificultad y discriminación del ejemplo propuesto

	Pretest (caso 1) **Grupo Sin Instrucción (caso 2)**		**Postest (caso 1)** **Grupo Con Instrucción (caso 2)**	
Sujetos (caso 1) **Sujetos emparejados (caso 2)**	**Ítem**		**Ítem**	
	1	**2**	**1**	**2**
Dificultad (pi)	0,4	0,2	0,7	0,8
DIS_{GN} (caso 1)	0,0 (3/10-3/10)	0,4		
$DIS_{CI\text{-}SI}$ (caso 2)	0,3 (7/10-4/10)	0,6		

3) Índice de homogeneidad. Tiene como propósito que los ítems que cubren un objetivo se comportan de forma similar, tanto en una aplicación del test como en sucesivas. Para medida de este índice se puede utilizar, como en los TRN visto en el capítulo 2, la correlación ítem-puntuación total del test.

3.3. Determinación de estándares y puntos de corte

En la corrección de una prueba criterial se necesita determinar una puntuación que sirva para diferenciar a los alumnos que han alcanzado o no el nivel mínimo de competencia exigible en la aplicación de la prueba. Esta puntuación determina el **estándar o punto de corte**. Aunque ambos términos se utilizan indistintamente debe saberse que el término estándar se refiere a una puntuación teórica, en tanto que punto de corte se utiliza para puntuaciones observables.

Para determinar los estándares o punto de corte se utiliza de nuevo el juicio de expertos. Estos individuos serán preferentemente los mismos que han definido el dominio e intervenido en la revisión lógica de los ítems propuestos.

Para cumplir sus objetivos los expertos habrán de asignar niveles de importancia o dificultad a los distintos objetivos en función de los niveles de logros de dichos objetivos por los individuos y de la relevancia que dicho objetivo tiene para la adquisición o aplicación en otras tareas.

Jornet y Suárez (1989b) plantean distintos métodos para la determinación de estándares y puntos de corte. Su exposición detallada se puede consultar en la referencia citada.

Tabla 8.4. Clasificación de los métodos de determinación de estándares y establecimiento del punto de corte (Jornet y Suárez 1989b, p. 279)

Tipo de objetivo métrico	Categorías metodológicas
Métodos para la determinación de un estándar	Métodos basados en juicios sobre ítems Métodos basados en juicios sobre los sujetos Métodos basados en el juicio sobre el grupo de referencia
Métodos para el establecimiento de un punto de corte	Modelos de estado Modelos continuos basados en la teoría de la decisión Métodos de compromiso Otros métodos

3.4. Estudio de las característica técnicas de la prueba

Comprende el estudio de la fiabilidad y validez de la prueba. La fiabilidad mide el grado de estabilidad de los resultados en aplicaciones sucesivas

de la prueba o de otra forma la fiabilidad de las clasificaciones establecidas mediante el test. La validez viene indicada por la adecuación de la prueba a los objetivos de la misma. Además el estudio de la validez debe atender a una triple dimensión: contenido, constructo y predictiva o de criterio.

Como coeficiente de fiabilidad se suele utilizar el índice de consistencia de distribuciones o índice kappa visto en el capítulo 2.

4. UN EJEMPLO DE PRUEBA CRITERIAL

Resulta difícil sintetiza en una pocas líneas todo el proceso inherente a la construcción de pruebas criteriales, no obstante el lector interesado puede consultar el texto: "Evaluación criterial del área de Matemáticas. Un modelo para la educación primaria" de Francisco Díaz Alcaraz y José Julián García García, editorial Praxis o el artículo "El diagnóstico de los Aprendizajes Básicos y sus Deficiencias. Un modelo criterial" de M.ª Isabel Salvador Pérez (consultas en Dialnet, diciembre 2010).

En Internet hay referencia a pruebas criteriales, fundamentalmente en idioma inglés: Universidad de Nevada (diversas pruebas de Matemáticas) o en la dirección http://dpi.state.wi.us/oea/pdf/math_framework.pdf (consulta realizada en diciembre 2010).

BIBLIOGRAFÍA

Berk, L. E. (1998). *Desarrollo del niño y del adolescente*. 4.ª ed. Madrid: Prentice Hall.

Jornet, J. M. y Suárez, J. M. (1989a). "Conceptualización del dominio educativo desde una perspectiva integradora en evaluación referida al criterio (ERC)". *Bordón 41(2)*. Págs. 237-275.

Jornet, J. M. y Suárez, J. M. (1989b). "Revisión de modelos y métodos en la determinación de estándares y en el establecimiento de un punto de corte en evaluación referida al criterio (ERC)". *Bordón 41(2)*. Págs. 277-301.

Kerlinger, F. N. (1985). *Investigación del comportamiento. Técnicas y metodología*. México: Interamericana.

De Lara, E. y Ballesteros, B. (2001). *Métodos de investigación en educación social.* Madrid: UNED.

López Feal, R. (1986). *Construcción de instrumentos de medida en Ciencias Conductuales y Sociales.* Barcelona: Alamex.

Martínez Arias, M. R. (1995). *Psicometría: Teoría de los Tests Psicológicos y Educativos*. Madrid: Síntesis.

Montoya Sáenz, J. M. (1988). *Actitudes sociopolíticas de los estudiantes españoles*. Madrid: UNED.

Morales Vallejo, P. (1988). *Medición de actitudes en Psicología y Educación*. San Sebastián: Tártalo.

Muñiz, J. (2000). *Teoría clásica de los tests.* Madrid: Pirámide.

Rincón, D. del; Arnal, J.; Latorre, A. y Sans, A. (1995). *Técnicas de investigación en Ciencias Sociales.* Madrid: Dykinson.

Sierra Bravo, R. (1985). *Técnicas de investigación social. Teoría y ejercicios*. Madrid: Paraninfo.

Silva, F. (1989). *Evaluación Conductual y Criterios Psicométricos*. Madrid: Pirámide.

Visauta, S. (1989). *Técnicas de investigación social*. Barcelona: PPU.

Walker, R. (1989). *Métodos de investigación para el profesorado.* Madrid: Morata.

Capítulo 9
TÉCNICAS GRUPALES DE RECOGIDA DE INFORMACIÓN

1. Introducción
2. Tipos
3. Grupos focales
4. Grupo nominal
5. Entrevistas en profundidad
6. El Brainstorming
   6.1. Requisitos de una sesión de brainstorming o tormenta de ideas
   6.2. Brainstorming paso a paso
7. El método Delphi
   7.1. Fases
8. Técnicas sociométricas
   8.1. El sociograma
   8.2. Escala de distancia social
   8.3. Los inventarios de personalidad

Objetivos

- Valorar la utilidad de técnicas grupales en las investigaciones sociales
- Conocer las distintas variantes de técnicas grupales y sus limitaciones
- Saber elegir el modelo de técnicas grupales más apropiado a un contexto concreto
- Diseñar las distintas técnicas grupales
- Realizar los controles técnicos apropiados para controlar la calidad de las técnicas grupales
- Saber implementar una técnica grupal en el contexto de una investigación
- Recoger y analizar los datos recogidos de una técnica grupal

1. INTRODUCCIÓN

Las técnicas grupales de recogida de información van dirigidas a identificar las conductas de las personas cuando se encuentran dentro de un grupo, formando la personalidad propia del grupo. Algunas de las técnicas vistas en capítulos anteriores, como la observación, se pueden concretar en los grupos dando como fruto una visión más holística por las diferentes aportaciones de cada uno de los miembros del mismo.

2. TIPOS

Aun cuando diferentes autores presentan distinta clasificación de las técnicas grupales de recogida de información, se muestra aquella de mayor consenso.

- Técnica de grupo focal
- Técnicas de grupo nominal
- Entrevista en profundidad
- Tormenta de ideas o Brainstorming
- Técnica delphi
- Técnicas sociométricas

3. GRUPOS FOCALES

Es una técnica en la que un moderador introduce un tema de discusión a un grupo de individuos, dándoles la oportunidad de que interactúen con sus comentarios y opiniones dirigidos en todo momento hacia los objetivos de análisis gracias a la habilidad del moderador.

En un proceso de evaluación se puede utilizar esta técnica para, por ejemplo:

a) **Obtener más información sobre percepciones, ideas, actitudes, experiencias, o creencias.** Los grupos focales son útiles para recopilación de perspectivas propias de partes interesadas en un proceso. Por ejemplo, una evaluación de un programa de prevención de la violencia en las aulas incluirá en los grupos focales a estudiantes de escuelas conflictivas. Los datos recogidos de los estudiantes serán acerca de las percepciones en relación a su participación, así como ese ambiente conflictivo afecta a su rendimiento escolar.

b) **Obtener información adicional como complemento a los datos cuantitativos recogidos.** Los grupos focales, al igual que otros métodos cualitativos, son útiles para proporcionar interpretaciones de los datos recogidos a través de métodos cuantitativos. Por ejemplo, sea un conjunto de profesores que recibe formación en la metodología de resolución de problemas como técnica formativa, después se les pasa un cuestionario que pretende recoger el nivel de conocimientos adquiridos. Un enfoque de grupo focal llevado a cabo durante varios meses más tarde podría proporcionar información adicional acerca de cómo los profesores están usando los nuevos conocimientos en su trabajo, lo que ellos creen respecto a los beneficios de la formación recibida, y qué ideas tienen acerca de cómo la formación se podría mejorar.

El proceso de aplicación de la técnica de grupo focal discurre por los siguientes hitos:

1. Elaboración de una *guía para el moderador* con un conjunto de preguntas y mensajes para llevar a buen término la reunión. El árbitro hará preguntas al grupo y dará tiempo para su contestación. La guía servirá de "hoja de ruta" para refrescar la memoria del moderador.

2. *Selección del número y tipo de participantes para cada grupo focal*. Según el conjunto de temas a tratar así buscará el moderador individuos similares para formar grupos homogéneos y obtener el máximo de información de cada tema.

3. Se necesita un facilitador o moderador, un encargado de tomar notas y un técnico.

a) El facilitador guía al grupo a través de la discusión y mantiene al grupo centrado en los temas de debate.

b) El encargado de tomar notas es un observador y no interactúa con el grupo. Las notas deben incluir las sensaciones de lo que cada persona dijo; identificar cómo observaciones lo que se dijo, y registrar cuando se realizaron las transiciones entre cada temática.

c) El técnico es responsable de la grabación del grupo focal. La grabación se utilizará para realizar una transcripción del evento. Naturalmente con los avances de los medios técnicos puede ser la misma persona que toma notas quien sea el responsable técnico de la grabación.

Ventajas de los grupos focales:

- Son relativamente fáciles de configurar.
- La dinámica de grupo puede proporcionar información que la recogida de datos individuales no proporciona.
- Es útil para el conocimiento más profundo de un tema, cuestión que puede ser más difícil obtener a través de otros métodos de recogida de datos.

Inconvenientes de los grupos focales:

- El facilitador es susceptible de caer en algún tipo de sesgo.
- La discusión grupal puede ser dominada o desviada por unos pocos individuos.
- El análisis de datos lleva mucho tiempo y necesita ser bien planificado con antelación.
- No proporciona información válida a nivel individual.
- La información recogida no es representativa para otros grupos.

4. GRUPO NOMINAL

Se denomina grupo nominal a una reunión grupal de grupo pequeño en la que se fomenta la participación de todos los miembros del grupo, evitan-

do la monopolización de la reunión por los dirigentes, sobre todo ante causas de inhibición como jerarquía, inseguridad, timidez... Tiene como objetivo producir un gran número de ideas en un corto periodo de tiempo.

El proceso de reunión se realiza de la siguiente manera: se plantea una pregunta al grupo de forma verbal o por escrito. Individualmente y en silencio, cada participante escribe sus ideas durante un tiempo prudencial de 10 minutos máximo. En una reunión plenaria, cada miembro defiende sus ideas durante 15 minutos máximo y se realiza una puesta en común. Las defensas de las ideas de cada uno se discuten en el grupo con un tratamiento de tiempo igualitario, que en votación se establece unas prioridades o se realiza un análisis de la situación generándose ideas innovadoras.

La técnica del grupo nominal se utiliza para:

— Generar ideas creativas
— Cuando existe la preocupación de que algún miembro del grupo no tiene la suficiente libertad de expresar sus ideas en otro contexto
— Para lograr consensos
— Cuando existe conflictos y divergencias en el grupo

5. ENTREVISTAS EN PROFUNDIDAD

La entrevista en profundidad, es el proceso de comunicación mediante interacción dinámica entre dos personas, entrevistador y entrevistado bajo control del primero.

La finalidad es conseguir información lo más relevante posible, sobre un objeto de análisis. Se suele utilizar cuando:

— se desea conseguir información muy compleja
— se busca información confidencial o delicada, por ejemplo sobre un tema tabú
— se busca información de profesionales, y los cuestionarios estructurados son insuficientes
— se considera como fase previa a la elaboración de cuestionarios estructurados, para identificar contenidos a incluir o incluso aspectos relevantes

Proceso de realización:

1. Elaboración del guión:
   — Longitud de la entrevista (corta con respuestas breves /larga)
   — Naturaleza de las preguntas (respuesta concisa, explicativa)
   — Naturaleza de la investigación (exploratoria, descriptiva, confirmatoria)
2. Fase introductoria:
   — Finalidad de la entrevista y uso de la información obtenida
   — Confidencialidad y anonimato de las respuestas
   — Tipo de colaboración deseada para el entrevistado
3. Desarrollo
   — No ser muy directo al preguntar y comenzar por aspectos poco comprometidos
   — Uso de antecedentes, o referencias a respuestas anteriores, para forzar en cierta medida la coherencia en las respuestas
   — Evitar preguntas tendenciosas o que incluyan un posible sesgo
   — Tener una secuencia lógica, de los genérico a lo específico, de los superficial a lo comprometido
   — Finalizar de forma distinguida y dejando la puerta abierta a otra entrevista

6. EL BRAINSTORMING

Es una técnica de creatividad desarrollada en una reunión de grupo con personas seleccionadas.

Ciertamente, los grupos no son imprescindibles para estimular el pensamiento creativo. **Aun individualmente es posible hacer el** *brainstorming*. Además, en un grupo, debemos escuchar a los otros y quizás utilizar tiempo para repetir nuestras ideas para hacerlas comprender a los demás.

La técnica conocida como brainstorming o tormenta de ideas puede ser una forma efectiva de generar cúmulos de ideas sobre un asunto específico para luego determinar que idea —o ideas— presenta la mejor solución.

El brainstorming resulta más provechoso cuando se hace en grupos de entre 8 y 12 personas y en un ambiente relajado.

Si los participantes se encuentran relajados, expandirán más sus mentes y producirán más ideas creativas.

6.1. Requisitos de una sesión de brainstorming o tormenta de ideas

Una sesión de brainstorming require de un facilitador, espacio libre y algo donde ir anotando las ideas, como un panel o varios papeles de gran tamaño o un software especial.

Las responsabilidades del facilitador incluyen: guiar la sesión, alentar la participación de todos los convocados y tomar notas de las ideas sin analizarlas o juzgarlas en esta etapa.

El Brainstorming funciona mejor si el grupo de personas es variado. Lo ideal es que los participantes pertenezcan a distintos departamentos de la organización y tengan diferentes perfiles profesionales. Aún en las áreas especializadas, los agentes externos pueden aportar ideas frescas e inspirar a los expertos.

Hay varios enfoques para una sesión de brainstorming, pero el enfoque tradicional es generalmente el más efectivo porque es el más enérgico y abiertamente colaborador, permitiendo que los participantes utilicen las ideas de los otros para proponer nuevas.

Los ejercicios creativos, los ejercicios de relajación y otras actividades entretenidas realizadas antes de comenzar, ayudarán a los participantes a relajar sus mentes para poder ser más creativos durante la sesión.

6.2. Brainstorming paso a paso

- **En primer lugar se debe definir el problema o asunto como un desafío creativo.** Es muy importante. Un desafío mal diseñado puede

llevar a producir un cúmulo de ideas que no sirvan para resolver el problema. En general, los desafíos creativos comienzan con las siguientes preguntas: ¿De qué manera podríamos...?, o ¿Cómo podríamos...?. El desafío debe ser conciso, directo al punto y excluir toda información no relacionada con el mismo. Por ejemplo: ¿De qué formas podríamos mejorar los resultados académicos del grupo X?

- **En segundo término se debe fijar un tiempo límite**. Una buena medida pueden ser 25 minutos, pero la experiencia con el grupo definirá el tiempo requerido. Los grupos más grandes suelen necesitar más tiempo para que todos puedan oportar ideas. También se puede fijar un límite de ideas. Un valor razonable son 50 ideas, pero se debe tener presente que cuanto más ideas surjan mucho mejor.
- Una vez comenzado el brainstorming, los participantes dan propuestas y el facilitador toma nota, usualmente en una pizarra o en un rotafolio para que todos las vean. De ninguna manera se deben criticar las ideas. No importa lo descabelladas o imposibles que parezcan, todas deben ser escritas en la pizarra. Se debe fomentar la risa, no la crítica.
- Una vez concluido el tiempo, **seleccionar las cinco ideas de mayor consenso**.
- **Ahora se deben escribir cinco criterios para juzgar cuales ideas solucionan mejor el problema o desafío**. Los criterios deberían empezar con la palabra "debería", por ejemplo: "Debería ser operativo", "debería poder completarse antes del 10 de abril", "debería ser económico", etc.
- **Dar a cada idea una valoración de 0 a 5 puntos** dependiendo del cumplimiento de los criterios anteriormente mencionados. Una vez que todas las ideas han sido evaluadas criterio por criterio, se suma la valoración total.
- **La idea con la valoración total más alta será la que mejor solucione su problema**. Pero se debe guardar un registro con las mejores ideas y su valoración, para el caso de que la idea seleccionada no sea realizable.

7. EL MÉTODO DELPHI

De forma sintética, el método Delphi se realiza mediante la distribución de una serie de cuestionarios en varias rondas a un grupo de expertos que puntúan sus opiniones en una escala cuantitativa. En cada ronda, los panelistas reciben realimentación sobre las puntuaciones del resto del grupo y tienen la oportunidad de revisar y cambiar su puntuación, si lo desean. El proceso termina cuando el investigador considera que se ha llegado a un determinado nivel de consenso o que la información obtenida es suficiente.

Entre otras aplicaciones, sirve de instrumento de análisis y compresión de realidades complejas, en las que es necesario contar con opiniones de individuos informados o implicados en las mismas (Landeta, 2002, p. 137).

El método Delphi, además, presenta ventajas respecto a otras técnicas grupales, especialmente frente a los grupos de discusión cara a cara. Entre ellas, la más destacable es que ningún miembro del grupo conoce las respuestas particulares del resto de participantes en el mismo, lo que reduce el efecto pernicioso que ejercen los miembros dominantes de un grupo en la opinión y respuestas de los demás, cuando éstas no son anónimas.

No obstante, el Delphi tradicional también tiene inconvenientes respecto a los grupos de discusión cara a cara. Uno de ellos es que el intercambio de información entre los participantes está limitado, al impedirles interactuar directamente. Adicionalmente, parece que reduce la motivación de los expertos, al no existir recompensa social por su contribución individual (Landeta, 2002, p. 157).

Este método presenta tres características fundamentales:

- Anonimato: Durante un Delphi, ningún experto conoce la identidad de los otros que componen el grupo de debate. Esto tiene una serie de aspectos positivos, como son:
  - — Impide la posibilidad de que un miembro del grupo sea influenciado por la reputación de otro de los miembros o por el peso que supone oponerse a la mayoría. La única influencia posible es la de la coherencia de los argumentos.
  - — Permite que un miembro pueda cambiar sus opiniones sin que eso suponga una pérdida de imagen.

— El experto puede defender sus argumentos con la tranquilidad que da saber que en caso de que sean erróneos, su equivocación no va a ser conocida por los otros expertos.

- Iteración y realimentación controlada: La iteración se consigue al presentar varias veces el mismo cuestionario. Como, además, se van presentando los resultados obtenidos con los cuestionarios anteriores, se consigue que los expertos vayan conociendo los distintos puntos de vista y puedan ir modificando su opinión si los argumentos presentados les parecen más apropiados que los suyos.
- Respuesta del grupo en forma estadística: La información que se presenta a los expertos no es sólo el punto de vista de la mayoría, sino que se presentan todas las opiniones indicando el grado de acuerdo que se ha obtenido.

En la realización de un Delphi aparece una terminología específica:

Circulación

Es cada uno de los sucesivos cuestionarios que se presenta al grupo de expertos.

Cuestionario

El cuestionario es el documento que se envía a los expertos. No es sólo un documento que contiene una lista de preguntas, sino que es el documento con el que se consigue que los expertos interactúen, ya que en él se presentarán los resultados de anteriores circulaciones.

Panel

Es el conjunto de expertos que toma parte en el Delphi.

Moderador

Es la persona responsable de recoger las respuestas del panel y preparar los cuestionarios.

7.1. Fases

Antes de iniciar un Delphi se realizan una serie de tareas previas:

- Delimitar el contexto y el horizonte temporal en el que se desea realizar la previsión sobre el tema en estudio.

- Seleccionar el panel de expertos y conseguir su compromiso de colaboración. Las personas que sean elegidas no sólo deben ser grandes conocedores del tema sobre el que se realiza el estudio, sino que deben presentar una pluralidad en sus planteamientos. Esta pluralidad evitará la aparición de tendencias en la información disponible en el panel.
- Explicar a los expertos en qué consiste el método. De esta manera se puede conseguir la obtención de previsiones fiables, pues los expertos van a conocer en todo momento cuál es el objetivo de la cada una de los procesos que requiere la metodología.

En un Delphi clásico se pueden distinguir cuatro circulaciones o fases:

- Primera circulación

  El primer cuestionario es desestructurado, no existe un guión prefijado, sino que se pide a los expertos que establezcan cuáles son los eventos y tendencias más importantes que van a suceder en el futuro en el ámbito estudiado.

  Cuando los cuestionarios son devueltos, éste realiza una labor de síntesis y selección, obteniéndose un conjunto manejable de eventos, en el que cada uno está definido de la forma más clara posible. Este conjunto formará el cuestionario de la segunda circulación.

- Segunda circulación

  Los expertos reciben el cuestionario con los sucesos y se les pregunta por la fecha de ocurrencia. Una vez contestados, los cuestionarios son devueltos al moderador, que realiza un análisis estadístico de las previsiones de cada evento. El análisis se centra en el cálculo de la mediana, primer y tercer cuartil donde piensan los expertos van a suceder los hechos previstos.

  El moderador confecciona el cuestionario de la tercera circulación que comprende la lista de eventos y los estadísticos calculados para cada evento.

- Tercera circulación

  Los expertos reciben el tercer cuestionario y se les solicita que realicen nuevas previsiones. Si se reafirman en su previsión anterior y ésta queda fuera de los márgenes entre los cuartiles inferior y superior,

deben dar una explicación del motivo por el que creen que su previsión es correcta y la del resto del panel no. Estos argumentos se realimentarán al panel en la siguiente circulación. Al ser estos comentarios anónimos, los expertos pueden expresarse con total libertad, no estando sometidos a los problemas que aparecen en las reuniones cara a cara.

Cuando el moderador recibe las respuestas, realiza de nuevo el análisis estadístico y, además, organiza los argumentos dados por los expertos cuyas previsiones se salen de los márgenes intercuartiles. El cuestionario de la cuarta circulación va a contener el análisis estadístico y el resumen de los argumentos.

- Cuarta circulación

Se solicita a los expertos que hagan nuevas previsiones, teniendo en cuenta las explicaciones dadas por los expertos. Se pide a todos los expertos que den su opinión en relación con las discrepancias que han surgido en el cuestionario. Cuando el moderador recibe los cuestionarios, realiza un nuevo análisis y sintetiza los argumentos utilizados por los expertos.

Teóricamente, ya habría terminado el Delphi, quedando tan sólo la elaboración de un informe en el que se indicarían las fechas calculadas a partir del análisis de las respuestas de los expertos y los comentarios realizados por los panelistas. Sin embargo, si no se hubiese llegado a un consenso, existiendo posturas muy distantes, el moderador debería confrontar los distintos argumentos para averiguar si se ha cometido algún error en el proceso.

Un ejemplo de aplicación de la técnica delphi se puede encontrar en: http://www.cesmurcia.es/estudios/estudios_22/04_Capitulo_03_Estudio_22.pdf (consulta 24-8-2010)

8. TÉCNICAS SOCIOMÉTRICAS

A diferencia de las técnicas de observación, cuya atención se concentra en el estudiante como individuo, las técnicas sociométricas concentran su atención en el grupo y las interacciones sociales que se presentan en él.

Entendemos por "interacciones sociales" todas las situaciones en que las actitudes de los miembros del grupo se encuentran entrelazadas de tal manera que influyen fuertemente sobre las conductas que cada miembro en lo personal habrá de manifestar. En otras palabras, hablamos de la repercusión que la situación del grupo tiene en cada uno de los miembros.

El test sociométrico pretende valorar las relaciones sociales de rechazo o atracción, preferencia, etc.

Los principales instrumentos que utilizan las técnicas sociométricas son: *el sociograma, la escala de distancia social y los inventarios de personalidad*.

8.1. El sociograma

El sociograma es una técnica que pretende poner de manifiesto los lazos de influencia y de preferencia que existen en el grupo. En educación es manifiesta su utilidad, por ejemplo, para un docente es de gran utilidad conocer la presencia de lazos de vínculo entre sus estudiantes con el fin de utilizarlos a favor de la promoción del grupo.

Como resultado de la aplicación de sociogramas, el docente puede:

- Detectar los alumnos que son rechazados por el grupo,
- Descubrir la presencia de bandos antagónicos dentro del grupo,
- Auxiliarse de los alumnos de mayor influencia para orientar positivamente al grupo, etc.

Por tanto puede, entre otras muchas utilidades:

- *Detección de problemas de inadaptación social*. Detectar la existencia de elementos aislados en el grupo, no elegidos por nadie, ante lo cuál habrá que procurar que los demás les preste atención. Detecta la existencia de un miembro rechazado por la mayoría de los miembros, ante lo cuál, será necesario examinar las causas que provocan esta marginación. También detecta la existencia de subgrupos ante lo cuál será conveniente hablar con los líderes.
- *Determinar estrategias metodológicas en el grupo*. Especialmente en los momentos previstos para la conformación de grupos para realizar

tareas, será un dato valioso el de la estructura social o informal del grupo en su conjunto. Se procurará, en principio, que los grupos de trabajo se ajusten a los grupos aparecidos espontáneamente, en función de las preferencias manifestadas por los miembros. Darán un resultado óptimo, siempre que el grupo esté bien integrado y sin problemas especiales. Cuando haya algún sujeto con dificultades de integración, el conocimiento estructural del grupo servirá para incorporarlo con otros miembros que le pueden motivar y con los que él se encuentre más compenetrado, tanto para el trabajo como para cualquier otro tipo de actividad. Resultará, así, más fácil romper el aislamiento.

Ejemplo 1. Si un profesor pretende conocer cuáles son los estudiantes con mayor influencia en el grupo, podría pedir a todos que contestaran por escrito la siguiente pregunta:

¿Cuáles de tus compañeros consideras que dan opiniones muy acertadas en la mayoría de los casos?

En primer lugar _

En segundo lugar _

En tercer lugar _

Ejemplo 2. Si un profesor pretende conocer cuáles son los "grupitos naturales" que por afinidad se forman dentro del grupo de clase, podría pedirles que contestaran lo siguiente:

¿Con cuáles de tus compañeros prefieres convivir ya sea en el estudio o en la diversión? ¿Y cuáles les rechazas?

En ambos ejemplos, una vez que el profesor cuenta con los datos, deberá hacer un minucioso análisis de ellos para obtener conclusiones. Una forma muy práctica de visualizar los resultados es representar gráficamente cada estudiante por medio de un círculo y de allí trazar flechas hacia los círculos de los compañeros que ha seleccionado. Por tanto, *los sociogramas se caracterizan por la representación gráfica de las elecciones y rechazos de los individuos de un grupo*.

Para realizar un sociograma se parte de una matriz de doble entrada llamada *matriz sociométrica* donde en filas, por ejemplo, se ponen los electores/as y en columnas los elegidos/as. Supongamos unos resultados ficticios

del ejemplo 2 donde se supone hay 13 estudiantes en el grupo y se da las siguientes elecciones (aceptación en azul, rechazo en rojo)

Tabla 9.1. Matriz sociométrica (ejemplo 2)

	1	2	3	4	5	6	7	8	9	10	11	12	13
1				**1**				1					
2	**1**						1						
3											1	**1**	
4			1		**1**								
5				**1**			1						
6			1							**1**			
7				**1**				1					
8						**1**			1				
9		**1**					1						
10					**1**						1		
11								1				**1**	
12		**1**					1						
13			1								**1**		

Se pueden representar los resultados del ejemplo 2 mediante el siguiente sociograma:

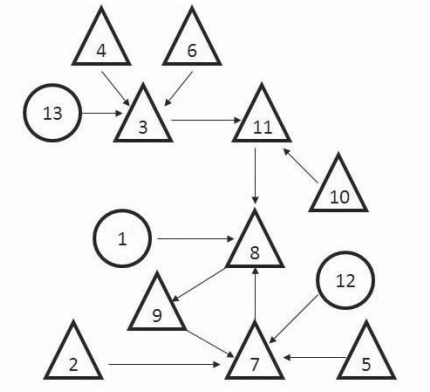


Figura 9.1a. Sociograma-aceptación.

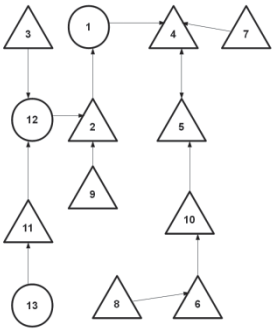


Figura 9.1b. Sociograma-rechazo.

Donde los triángulos representan a los alumnos y los círculos a las alumnas y la flecha representa elección (parte del elector hacia el elegido).

A la hora de interpretar el sociograma hay dos cuestiones a tener en cuenta: a) la posición sociométrica y b) las configuraciones sociométricas.

Existen distintas tipologías de posiciones sociométricas:

1) **El líder**. Una persona que recibe muchas elecciones y pocos rechazos. Goza, del mayor índice de popularidad y status sociométrico. En el ejemplo el líder será el alumno 7. Puede no existir un solo líder que influya sobre todos los miembros del grupo; es más probable que el grupo esté dividido en dos o más subgrupos enfrentados entre sí y que cada uno de ellos tenga su propio líder. Cuando en el proceso de análisis se localizan subgrupos es importante ver si existen individuos que actúan como puente entre uno y otro, y luego valorar la relación de este sujeto "puente" con el líder, etc. Las relaciones entre los distintos líderes son importantes para comprender la dinámica del grupo; así, si los distintos líderes están comunicados entre sí por relaciones bidireccionales puede hablarse de un autoritarismo de esos individuos que controlan y dominan el grupo.

2) **La eminencia gris.** Persona prácticamente aislada, que casi sólo posee una relación mutua con el líder. En el ejemplo hay tres candidatos a dicho puesto, los estudiantes 2, 5 y 12. Su influencia grupal es muy grande, aunque externamente no se note, al ser su popularidad muy baja.

3) **El aislado-a.** Persona que no recibe elecciones y tampoco elige a nadie Para detectarlo es necesario que el sujeto tenga entera libertad para elegir o dejar en blanco las preguntas. En nuestro caso no se detecta ningún estudiante aislado.

4) **El marginado-a.** Individuo que elige a otros, pero que no es elegido por nadie. En el ejemplo serán marginados/as todos los estudiantes excepto los individuos: 3, 7, 8, 9 y 11. Tanto los individuos aislados como los marginados no suelen presentar problemas al profesor, sin embargo deben ser considerados por el mismo al existir una problemática soterrada que les lleva a ese posicionamiento. Se aconseja estudiarlos con detenimiento, y tratar de integrarlos en el grupo, asig-

nándole, al comienzo, tareas fáciles que les hagan triunfar ante los demás y que, por tanto, poco a poco sean tenidos en cuenta.

5) **El rechazado-a parcial**. Persona que recibe menos aceptaciones que rechazos. En nuestro ejemplo hay varios individuos con estas características: 1, 6, 10, 2, 5, 12 y 4 . Su presencia suele notarse en el grupo, ya que normalmente es bastante bullicioso. Suele hacerse el gracioso.

6) **El rechazado-a total**. Sujeto que solamente recibe rechazos. En el ejemplo es el individuo 4 el que más rechazos recibe aunque 1, 6, 10, 2, 5 y 12 también reciben sólo rechazos. Su presencia suele resultar pesada tanto para el grupo como para el profesor. Presenta muchas características negativas para el grupo. Las causas suelen encontrarse en problemas familiares y/o sociales de su entorno. Estos sujetos presentan graves problemas de disciplina. Fracasan con frecuencia en los estudios. Suelen faltar a clase con cierta frecuencia.

7) **El normal**. Persona que recibe algunas aceptaciones y algunos rechazos. Suelen ser la mayoría de los miembros del grupo. Tal sería el caso del individuo 11.

8) **El polémico-a**. Es el sujeto que recibe bastantes aceptaciones y bastantes rechazos. En el ejemplo al ser pocos individuos no se puede notar muy bien dicha característica, el individuo 11 sería el mayor candidato al puesto. Presenta a la vez características positivas y negativas. Suele dividir al grupo y se presenta normalmente como el "antilíder" dentro del grupo, y su influencia y prestigio social es bastante elevada.

Respecto a las configuraciones sociométricas existen también distintas variantes:

1) **La pareja o relación diádica**. Dos individuos A y B se eligen mutuamente. En el ejemplo propuesto no existe ese caso.

2) **El triángulo o relación triádica**. A, B y C se eligen entre sí. En nuestro ejemplo se puede considerar al grupo 7, 8 y 9 si se considera a uno de ellos como puente. Suelen formar un subgrupo muy unido y compacto. Su influencia en la dinámica del grupo suele ser grande si uno de ellos sirve de individuo puente. La existencia de muchos subgrupos nos permite concluir que el grupo está disgregado, poco cohesionado,

y normalmente están enfrentados entre sí, al no entenderse sus líderes respectivos.

3) **La cadena**. A elige a B, B a C, C a D...(puede haber elecciones recíprocas, aunque no son necesarias para la cadena). En el ejemplo hay varias cadenas una por ejemplo es la formada por los individuos 1, 8, 9 y 7. Cuanto más próximo esté un sujeto al líder, mayor será su status sociométrico.

4) **La estrella**. Se da este tipo cuando un sujeto recibe muchas flechas. Refleja la figura del líder. Sirve de eje al grupo y, de ahí, su influencia en la dinámica del grupo es destacada. Su colaboración o no con el profesor va a determinar que el grupo acepte o rechace al profesor. Es necesario encomendarle tareas difíciles y complejas. Puede ayudar al docente a ser respetado por la clase.

8.2. Escala de distancia social

En los grupos de clase hay estudiantes que tienen buenas relaciones con la mayoría de sus compañeros; otros se relacionan sólo con un grupo reducido; hay también quienes no se relacionan, bien por el rechazo de sus compañeros, bien por voluntad propia. También hay grupos que reaccionan de forma unánime ante alguna propuesta de los profesores y otros donde las opiniones son dispares al existir grupos a menudo enfrentados y casi nunca una propuesta recibe el apoyo o el rechazo por unanimidad.

Para comprobar la existencia de estas situaciones se pueden utilizar los llamados **índices de integración y cohesión**. Se puede decir que un sujeto está "integrado" en su grupo cuando mantiene una red de interrelaciones con otros miembros del grupo, que le generan la sensación de dependencia al mismo. Entonces se definirá la integración como una relación particular del sujeto respecto al grupo. Los resultados del índice podrán compararse con los del test sociométrico con los que deben tener una estrecha relación.

Por otra parte diremos que un grupo está "cohesionado" cuando el sistema de interrelaciones que funciona entre sus miembros, mantiene a estos satisfactoriamente ligados al grupo. El concepto de cohesión no es, por tanto, una cuestión de cada miembro del grupo sino algo inherente al grupo en su conjunto.

Las escalas de distancia social nos permiten conocer la opinión y actitud afectiva de cada estudiante respecto de sus compañeros. Para dicho objetivo se redactan proposiciones que van desde la mayor intimidad a la mayor lejanía, posteriormente cada estudiante enjuicia la situación en que se encuentra cada compañero respecto a él.

Veamos un ejemplo:

Sean 7 alumnas (Ana, Luisa, Juana, María, Mónica, Eva y Sandra) que contestan a la siguiente pregunta: "valorar el interés que tienes en que tu compañera ..... pertenezca a tu equipo". Donde las valoraciones son:

5..... Tengo mucho interés en que figure en mi equipo

4..... Prefiero que esté en mi equipo

3..... Me es indiferente

2..... Prefiero que no forme parte de mi equipo

1..... Quisiera no tener ningún tipo de contacto con ella

Los resultados de este ejemplo simulado son los siguientes:

	Valoración de cada estudiante respecto a las demás							
	Ana	Luisa	Juana	María	Mónica	Eva	Sandra	Total
Ana	——	4	2	1	5	1	2	15
Luisa	3	——	4	5	2	2	4	20
Juana	3	1	——	3	3	1	4	15
María	4	3	3	——	5	4	4	23
Mónica	3	2	4	3	——	3	5	20
Eva	1	2	4	2	2	——	3	14
Sandra	2	2	4	2	5	1	——	16

La escala, para situar a cada alumno según su *índice de integración,* oscilará entre 5(n-1) y (n-1), siendo n=7 (en este caso). Es decir entre 30 y 6 respectivamente.

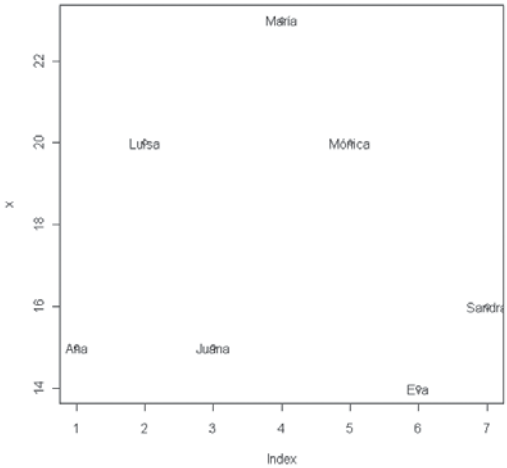


Figura 9.2. Representación gráfica de los índices de integración.

Según se puede observar la estudiante con mayor integración en el grupo es María y la de menor grado de integración es Eva.

A la vista de los resultados, el profesor u orientador escolar podrá ayudar a los estudiantes con bajo índice de integración para que consigan mejorar las relaciones con los demás miembros del grupo. A través del test sociométrico se puede saber si el problema de integración se produce por rechazo de sus compañeros o por falta de relaciones con ellos.

Para calcular la cohesión se puede utilizar el índice de cohesión que es la media de los índices de integración de los estudiantes del grupo. Nos da una información muy valiosa y de conjunto sobre un grupo determinado, en comparación con otros grupos o el mismo grupo después de un cierto tiempo.

En general:

$$\text{Índice de cohesión} = \frac{\sum \text{índices de integración}}{\text{Número de miembros}}$$

En el ejemplo propuesto:

$$\text{Índice de cohesión} = \frac{15+20+15+23+20+14+16}{7} = 17{,}57$$

Se valora con la escala que ya hemos visto de 30 máximo y 6 mínimo. Como vemos un valor muy en el centro de la escala.

El índice de cohesión también se debe estudiar en relación con los resultados del test sociométrico. Se comprobará la existencia de una gran dispersión en las relaciones intragrupales o la existencia de subgrupos más o menos conformados, fenómeno este último de lo más natural.

Un ejemplo de análisis de redes sociales, lo puede encontrar el lector en: http://www.griss.org/curriculums/fernandezquijada/publicaciones/aeic_fernandezquijada.pdf (consulta 24-8-2010). Además existe software específico de análisis de redes sociales, que utiliza indicadores relativos a las distancias entre nodos de la red, indicadores sociométricos y medidas de centralización. Dentro de este software destacan programas como Ucinet y Agna, este último, software libre (http://www.freewebs.com/benta/agna/index.htm).

8.3. Los inventarios de personalidad

Los inventarios de personalidad consisten en poner al estudiante en una situación ficticia en relación a una conducta social, pero a la que muy posiblemente puede tener acceso, para que explique qué haría o dejaría de hacer en esas circunstancias; todo esto con el fin de que proyecte sus actitudes en relación con conductas sociales específicas.

Sea por ejemplo el siguiente caso:

El profesor plantea a los estudiantes la siguiente situación: en una clase se han formado dos grupos; el primero de ellos se ha integrado por los estudiantes más agresivos que quieren sentirse dominadores del grupo y atemorizar a los demás; en el segundo, se ha integrado el resto de los estudiantes. Este segundo grupo no quiere responder cada estudiante de forma aislada ante la situación reinante, pero no saben cómo actuar, no tienen un “jefe” o “líder” que los organice.

Explica detalladamente: ¿qué harías tú, si pertenecieras al segundo grupo y se te nombrara como jefe del mismo?

En las diferentes respuestas que los estudiantes señalen, el profesor tendrá material suficiente para valorar la actitud de los mismos hacia las situaciones en que todo el grupo está implicado.

BIBLIOGRAFÍA

Bisquerra, R. (1989). *Métodos de Investigación educativa. Guía práctica*. Barcelona: Ceac.

Bunge, M. (1976). *La investigación científica*. Barcelona: Ariel.

Calvo, C. (2007). *Técnicas e instrumentos de diagnóstico en educación.* Málaga: Aljibe.

Corbetta, Piergiorgio (2007). *Metodología y técnicas de investigación social.* Madrid: McGraw-Hill / Interamericana de España.

De Ketele, J. M. y Roegiers, X. (1995). *Metodología para la recogida de información*. Madrid: La Muralla.

Del Rincón, D. y otros (1995). *Técnicas de investigación en Ciencias Sociales*. Madrid: Dykinson.

Delgado, J. M. y Gutiérrez, J. (1995). *Métodos y técnicas cualitativas de investigación en ciencias sociales.* Madrid: Síntesis.

Duverger, M. (1974). *Método de las ciencias sociales*. Barcelona: Ariel.

Fontana, A. y Frey, J. H. (1994). Interviewing: The Art of Science. En N. K. Denzin y Y. S. Lincoln (Eds.), *Handbook of Qualitative Research,* (pp. 361-376). Londres: Sage.

Ibáñez, J. (1979). *Más allá de la Sociología.* El grupo de discusión: Técnica y Crítica. Madrid: Siglo XXI. (pp.259-317).

Landeta, J. (1999). *El método delphi: una técnica de previsión para la incertidumbre.* Barcelona: Ariel.

Morales, J. F. (coord.) (1994). *Psicología Social*. Madrid: McGraw-Hill.

Kerlinger, F. N. (1985). *Investigación del comportamiento. Técnicas y metodología*. México: Interamericana.

Krueger, R. A. (1988) *El grupo de discusión. Guía práctica para la investigación aplicada.* Madrid: Pirámide.

Ruiz Olabuénaga, J. I. e Ispizúa, M. A. (1989) *La descodificación de la vida cotidiana. Métodos de investigación cualitativa.* Bilbao: Universidad de Deusto.

Simon, P. y Albert, L. (1983). *Las relaciones interpersonales*. (Ejercicios prácticos). Barcelona: Herder.

Valles, M. J. (1997). *Técnicas cualitativas de investigación social. Reflexión metodológica y práctica profesional.* Madrid: Síntesis.

Walker, R. (1989). *Métodos de investigación para el profesorado.* Madrid: Morata.

Capítulo 10
ANÁLISIS DE DOCUMENTOS

1. Análisis de documentos
2. Análisis de contenido
3. Tipos de documentos
4. Procedimientos a seguir en el análisis de contenido
5. Tipología de análisis de contenido
   5.1. Análisis de contenido temático
   5.2. Análisis de contenido semántico
   5.3. Análisis de contenido de redes
6. Ejemplo de análisis de contenido

Objetivos

- Valorar la utilidad del análisis de documentos en las investigaciones sociales
- Conocer las distintas variantes del análisis de documentos y sus limitaciones
- Saber elegir el modelo de análisis documental más apropiado a un contexto concreto
- Diseñar un análisis de documentos
- Realizar los controles técnicos apropiados para controlar la calidad del análisis de documentos
- Saber implementar un análisis documental en el contexto de una investigación
- Recoger y analizar los datos recogidos de un análisis documental

1. ANÁLISIS DE DOCUMENTOS

El análisis de documentos es la base de la investigación histórica y el fundamento de toda investigación que pretenda estar incluida en la realidad del momento en el que se desarrolla. Además, ***es el primer paso a dar después de plantearse el problema*** y todos los autores que han escrito sobre procesos de investigación así lo recomiendan.

Existen distintas formas de abordar el estudio de esta importante fuente de información:

a) Los métodos clásicos: se reducen a buscar información en el momento de la lectura, derivan de la crítica literaria y de la crítica histórica.

Se procede mediante dos tipos de análisis: el interno, por medio del cual se conoce el contenido; y el método externo, que se ha de encargar de estudiar el contexto del que se procede, a fin de precisar el grado de veracidad y su resonancia.

Existen diversos métodos clásicos, pero el más común es el ***método histórico***. Conviene tener en cuenta dos aspectos importantes: la autenticidad y la veracidad del documento.

b) Otros métodos son los cuantitativos. Para Duverger (1974) el ***método cuantitativo*** de análisis de documentos no pretende sustituir a los métodos clásicos. Con este método se intenta eliminar el subjetivismo de la persona que analiza los documentos. No se suele entrar en estudios en profundidad de la temática abordada. En la actualidad es uno de los campos en el que el uso de los ordenadores le presta una valiosísima ayuda.

Algunos de los métodos cuantitativos de análisis de documentos son:

1) La semántica cuantitativa

2) El análisis de contenido

3) Documentación bibliográfica

La semántica cuantitativa trata del estudio del vocabulario, del estilo y los modos de expresión. Se basa en el conteo de las palabras de un texto, clasificándolas según su uso en el lenguaje (verbo, complemento, etc) y valiéndose de procedimientos matemáticos calcular la asociación de palabras para analizar de esta forma el estilo del autor.

Al ***análisis de contenido***, por su relevancia, dedicaremos el resto del capítulo, por tanto aquí sólo diremos que es un procedimiento de recogida de datos que goza de todas las características propias de objetividad y relevancia, características del método científico.

Finalmente en la documentación bibliográfica, se incide en que el autor de un trabajo puede utilizar sus propias palabras para explicar un dato o para reflejar el juicio de un experto, o puede también citar las palabras textuales de la obra y de los críticos sobre la materia. Con los datos y las citas se incluye la documentación pertinente a cada una. Documentar un trabajo quiere decir incluir en él notas que indiquen exactamente de dónde se tomaron los datos y las citas que forman parte del escrito. La documentación bibliográfica de un trabajo contiene las notas necesarias y bibliografía pertinente.

2. ANÁLISIS DE CONTENIDO

Para Berelson (1952, p. 18) es "una técnica de investigación para la descripción objetiva, sistemática y cuantitativa del contenido manifiesto de la comunicación". Posteriormente se ha aplicado también al contenido latente, es decir, al no evidenciado. En síntesis se trata de transformar un documento escrito en datos cuantitativos.

La definición anterior lleva inherente ciertas condiciones del análisis de contenido:

— **Objetividad**, es decir, el proceso de análisis puede ser replicado por cualquier investigador llegándose a los mismos resultados.

— **Sistematización**, debe realizarse según pautas ordenadas llegando a la totalidad del contenido recogido.

— **Cuantificable**, para hacer viable procesos estadísticos de tratamiento de datos.

— **Manifiesto**, es decir, con evidencia del mismo. No se incluyen, en principio, el mensaje no manifiesto o soterrado de la comunicación.

Por otra parte Krippendorff (1990, p. 28), define el análisis de contenido como "la técnica destinada a formular, a partir de ciertos datos, inferencias reproducibles y válidas que puedan aplicarse a un contexto".

Para su aplicación se suelen recorrer los siguientes pasos:

1. Decidir si se va analizar todo el documento o una muestra. En este último caso seleccionarla
2. Definir la unidad de análisis: palabras, frases, oración, párrafo, temas, etc.
3. Categorización: identificar, seleccionar y estructurar las categorías del análisis
4. Fiabilidad: correlación entre las categorizaciones de varios codificadores
5. Análisis estadístico: frecuencias, tablas de contingencias, medidas de asociación, análisis discriminante, análisis de cluster, análisis de correspondencias, etc.

El propósito del análisis de contenido puede ser, por ejemplo: describir tendencias, analizar técnicas persuasivas, analizar el estilo, relacionar características, relacionar atributos, etc.

El análisis de contenido presenta unas ventajas frente a otras técnicas de recogida de datos y algún inconveniente.

Ventajas

- Técnica no intrusiva
- Información de bajo coste

Inconvenientes

- Escritos con diferentes propósitos, que muchas veces no coinciden exactamente con el propósito de la investigación
- Disposición de los documentos limitada

3. TIPOS DE DOCUMENTOS

Los tipos de documentos son muy variados:

- Actas, circulares
- Informes, programas
- Cartas, diarios, discursos
- Revistas, periódicos

4. PROCEDIMIENTO A SEGUIR EN EL ANÁLISIS DE CONTENIDO

Los pasos a realizar en la recogida de información del contenido de un documento son los siguientes:

1) Objetivo y contexto

Se debe determinar los *objetivos* perseguidos con la investigación, el *universo* objeto de estudio y la *elección del tipo de documento* y las *unidades de contenido* dentro del documento.

2) Unidades de análisis

Son los objetos de estudio para realizar con ellos clasificaciones o recuentos. Se pueden clasificar en: *unidades de muestreo* —porciones del universo observado que serán objeto de estudio—, *unidades de contexto* —sirve para dar significado a las unidades de registro— y *unidades de registro* —elemento primario del texto que hace referencia a una categoría—. Las unidades de registro pueden ser de distintos tipos:

a) léxicas

— términos
— palabras

b) proposicionales

— oraciones
— proposiciones

c) narrativas

— acciones de personajes en un tiempo y en un espacio

d) temáticas

— conceptos

— referencias

— valoraciones

e) argumentativas

— fruto de enunciaciones

f) discursivas

— fruto de discursos de sujetos

3) Reglas de numeración o recuento

Existen numerosas variantes:

- Presencia o ausencia de un código
- Frecuencia de aparición de un código
- Frecuencia ponderada, según la importancia del código, se le da un peso a su aparición
- Orden de aparición de los códigos
- Densidad de un texto. Es igual al sumatorio de las frecuencias de los códigos dividido por la suma total de todas las palabras. Si la expresión es porcentual habrá que multiplicar por cien.
- Nivel de concentración. Número de códigos distintos divididos por el sumatorio de las frecuencias de los códigos. Si la expresión es porcentual habrá que multiplicar por cien.

4) Categorización

Es la clasificación de los elementos del texto a partir de unos criterios previamente establecidos o por procedimientos lingüísticos. Este procedimiento históricamente ha sido el "cuello de botella" del tratamiento informático-estadístico del análisis de contenido, en la actualidad existe procedimientos automáticos mediante la ayuda de ordenador (por ejemplo programa PASW Text Analytics) de realizar dicho proceso aunque debe ser después el analista quien finalmente marque la categorización definitiva siguiendo los intereses de investigación.

Las categorías deben cumplir las condiciones de:

— exclusión mutua, es decir, no puede haber dos categorías conexas

— homogeneidad, para el contenido que refleja la categoría

— productividad, pensado en proporcionar índices para inferencias

— significativas, es decir, relevante, por ejemplo no tiene sentido categorías para preposiciones, conjunciones, etc.

— replicables, en el sentido de poder realizar cualquier investigador la misma categorización

A la hora de unir las categorías en serie se debe utilizar un criterio único. Por ejemplo aquellas categorías cuya frecuencia de aparición sea mayor que un valor dado.

5) Codificación

En el tratamiento estadístico de las categorías es relevante la asignación de códigos a cada una de las mismas, de esta manera se facilita el tratamiento de datos.

Naturalmente la codificación debe permitir que las categorías a las que representan cumplan las condiciones de exclusión y consistencia.

6) Fiabilidad y validez

En el análisis de contenido, la fiabilidad se puede estimar calculando el porcentaje de veces que varios codificadores independientes coinciden cuando codifican el mismo material. Para realizar el estudio de la misma se acostumbra, cuando menos, que sean dos codificadores los que realicen el proceso.

La validez vendrá reflejada por la **exhaustividad** cuando las categorías agotan todo el contenido del documento y por la **productividad** o capacidad de que el conjunto de categorías proporcione resultados ricos para plantear inferencias e hipótesis.

7) Análisis de los datos

Para analizar los datos recogidos del análisis de contenido será necesario:

a) Una fase descriptiva donde se analizará la frecuencia de aparición de las categorías y sus relaciones internas o externas con otras variables.

b) Una fase inferencial donde se intentará sacar conclusiones, es decir, ver el poder de generalización.

c) Una fase multivariante donde se analizará las estructuras complejas de las categorías y las relaciones de los bloques de contenido con éstas.

5. TIPOLOGÍAS DE ANÁLISIS DE CONTENIDO

Se suelen utilizar tres tipologías de análisis de contenido: el análisis de contenido temático, el análisis de contenido semántico y el análisis de redes.

5.1. Análisis de contenido temático

Se centra en la presencia de términos o conceptos con independencia de las relaciones entre los mismos. Se utiliza como técnica el recuento, la identificación y clasificación temática y la búsqueda de palabras en contexto. Como busca una temática en un contexto previamente se ha de definir la misma antes del comienzo del análisis.

El **recuento de palabras** resulta muy simple y desde luego muy fácil de realizar por procedimientos informáticos, no obstante será necesario tener en cuenta algunas cuestiones morfológicas: singular o plural, preposiciones, conjunciones, artículos, etc., unas para evitar duplicidad de significado y otras por no aportar información relevante.

La **identificación y clasificación temática**, se puede efectuar después de conocer la frecuencia de aparición de palabras, llevará a fijar términos relevantes para el objetivo de investigación. Por ejemplo en el discurso de un político[1] recogido de varias intervenciones del mismo podemos acopiar los siguientes términos:

[1] Ejemplo recogido de R. Colle (www.colle.cl)

comercio, comercial, comerciales, comerciante
concertación
contaminación
crear, creación, crea, creadora, creando, creatividad
déficit, deficitarias, deficientes
desafío
desarrollo, desarrollar, desarrollan, desarrollaremos, desarrollarse
descentralización, descentralizadas, descentralizado, descentralizados, descentralizar
economía, económica, económicas, económico, económicos
educación, educacional, educar, educativos, capacitación,
capacitarse, enseñanza, profesorado, profesor
empresa, empresas, empresarial, empresariales, empresario, empresarios
estado, estados, estatal, estatales
familia, familiar, familias
futuro, futuro
gasto, gastar
gobierno, gobiernos, gobernante, gobernar, gobierna
hombre, hombres
infraestructura
inversión, inversiones, invertir, inversionista, invertirán, inviertan, invierten, invirtiendo
jóvenes, joven, joven
ley, legal, legalmente, legislación
modernización, moderna, moderniza, modernizada, modernizar, modernizaremos, moderno
mujer, mujeres
municipal, municipales, municipalidad, municipios
necesario, necesaria, necesarias, necesarios, necesidad, necesidades, necesita
oportunidad, oportunidades
pobreza, pobres, necesitados
política, políticas, político, políticos
problema, problemas
progreso, progresos, progresar
recursos
salud
social, sociales

trabajo, trabajos, trabaja, trabajando trabajador, trabajadora, trabajadores vivienda, viviendas, habitacional

Otra forma será trabajar con una "lista abierta" de temas, para agregar descriptores a medida que aparecen nuevos temas en un corpus (todo el conjunto de texto del discurso) bajo estudio. Esta operación se puede realizar de forma manual o mediante aplicaciones (software, como por ejemplo el mencionado Pasw Text de SPSS) destinadas al análisis temático automático. Estos programas se apoyan en un "diccionario" donde se registran todos los términos a tener en cuenta así como los sinónimos relacionados, para de esta forma poder "extraer" los temas tratados y su frecuencia. Para el éxito del proceso es crucial la calidad de este diccionario. No obstante, existe la dificultad de interpretación automática de palabras homógrafas, para lo cual estos programas también han desarrollado procedimientos que eligen el significado en función del contexto.

Las palabras en contexto tratan de seleccionar palabras no simplemente por la frecuencia de aparición sino en función del contexto. Programas como el Anatex (sólo ejecutable en Linux) desarrollado por Raymond Colle (1988) permiten la creación de palabras registradas junto con su contexto (KWIC:Key-word in context), de esta forma al aparecer la palabra se seleccionará el contexto más apropiado y facilita la creación de núcleos de contenido. La técnica KWIC permite la búsqueda de cadena de caracteres de diferente amplitud desde raíces (comunes a varias palabras) hasta palabras compuestas o incluso frases completas.

5.2. Análisis de contenido semántico

El análisis semántico pretende sobre todo estudiar las relaciones entre los temas tratados en un texto. En consecuencia se han de definir los patrones de relaciones que se tomarán en cuenta: relaciones gramaticales o relaciones más propiamente semánticas, por ejemplo: "expresiones de posiciones favorables a una u otra decisión".

Existen varias técnicas de análisis semántico: análisis de matriz gramatical, análisis de matriz semántica, análisis de componentes, análisis valo-

rativo, análisis de núcleos y análisis actancial. Nos centraremos en el segundo por ser el de mayor utilización.

El análisis de "matriz semántica" exige como paso previo, se definan los modelos de relaciones que podrán ser codificados, esto nos lleva a elaborar la llamada "gramática semántica" o conjunto de modelos con la forma sintáctica de las relaciones semánticas que interesen. Por ejemplo: "expresión de problema - solución sugerida", para estudiar las demandas de los ciudadanos ante un gobierno o programa electoral, que implica una estructura (matriz) binaria del tipo Verbo (acción a realizar = solución) vs. Objeto (problema al cual corresponde). En el caso de problemas laborales, por ejemplo, la estructura sería tríadica: Sujeto (quién) - Verbo (acción pedida) - Objeto (problema), o cuaternaria: Demandante/sujeto - Acción/verbo - Demandado/implicado - Problema/objeto (si se estudia, por ejemplo, el tema de las demandas sociales presentadas en un medio de comunicación).

Los ejemplos anteriores nos llevan a la siguiente conclusión: la matriz definida depende directamente de la estructura del fenómeno estudiado y además debe permitir recoger las diversas formas de expresión del mismo, para lo cual es importante comprender la intención del hablante y tener en cuenta el contexto. La unidad de análisis es la clausula (fragmento del texto) que "encaja" en la matriz definida (y no se tiene en cuenta todo el resto del texto).

Existe software (en inglés) capaz de reconocer algunas de estas estructuras, como el KEDS (Kansas Events Data System).

5.3. Análisis de contenido de redes

El análisis de contenido de redes se centra en la ubicación relativa de ciertos componentes. Se pueden realizar análisis más complejos partiendo de un análisis de matriz semántica, teniendo en cuenta el tipo de relaciones entre los términos, conceptos o temas. Así, por ejemplo, existen algunos trabajos que han construido secuencias de relaciones de causa a efecto (donde es posible calcular el "peso" propio de cada causa en relación a todas las otras). También es posible construir redes a partir de estructuras gramaticales o semánticas (como el modelo Sujeto-Verbo-Objeto).

Un ejemplo de herramienta para la creación de redes semánticas es SemNet (sólo ejecutable en Mac). Es un software "de propósito general" en materia de representación de redes semánticas ya que sirve tanto para el análisis de contenido como para organizar ideas, representar estructuras de conocimiento u orientar el estudio (aplicaciones educacionales de tipo hipertextual). Las tipologías sugeridas por SemNet para la realización de redes semánticas son:

Relaciones de inclusión

- contiene/es contenido
- igual a/igual a
- tiene por parte/es parte de

Relaciones de característica

- ocurre en/es donde ocurre
- tiene por característica/es característica de
- tiene por función/es función de
- tiene por tamaño/es tamaño de

Relaciones de rol

- es agente/tiene por agente
- es input de/tiene por input
- es output de/tiene por output
- es objeto de/tiene por objeto
- informa/es informado por

Relaciones temporales

- desarrolla/es desarrollado por
- experimenta con/es investigado por
- sigue/precede

Relaciones activas

- absorbe/es absorbido
- actúa sobre/es objeto de acción de
- adquiere/pierde

- cambia de/cambia en
- causa/es causado por
- consume/es consumido por
- designa/es designado por
- emite/recibe
- envuelve/está envuelto por
- es origen/es originado por
- especifica/es especificado por
- forma/ es formado por
- modifica/es modificado por
- necesita/es una necesidad de
- provee/está provisto por
- proviene de/conduce a
- se convierte en/es convertido a partir de
- se junta con/se junta con
- sintetiza/es sintetizado por
- usa/es usado por

6. EJEMPLO DE ANÁLISIS DE CONTENIDO

Existe multitud de ejemplos que se pueden plantear. El lector seguro puede sugerirse varios casos fruto de cualquier tratamiento de las preguntas abiertas de cuestionarios o de transcripciones de reuniones de grupo. No obstante a continuación se propone un ejemplo y se presentan algunas de las múltiples posibilidades de análisis que se pueden realizar en el mismo.

En primer lugar se debe definir el problema de investigación, en nuestro caso será: ¿Cómo enseñar a nuestro hijo el gusto por las Matemáticas?¿Qué debemos hacer?

Las unidades de análisis serán los párrafos del texto de un informe realizado por un experto de relevancia internacional en la enseñanza de las Matemáticas.

Las unidades de registro serán las palabras de las cuales anotaremos su frecuencia de aparición en cada uno de los 7 párrafos del texto.

El proceso de categorización se realizará con el paquete **tm** de R. Para realizar dicho proceso será necesario cargar el mencionado paquete según se indica a continuación:

1. Se prueba si el paquete está cargado en las opciones por defecto de R, para lo cual en la pantalla RGui se selecciona: *Paquetes>Cargar paquete.* Si **tm** está en la lista que presenta, se concluye el proceso. Si, como es el caso, no está entre los paquetes cargados por defecto será necesario instalar dicho software.

2. Para instalar un paquete en RGui se selecciona: *Paquetes>Instalar paquete*. Aparece un CRAN mirror que son servidores del programa R en diferentes sitios del mundo, nosotros por proximidad seleccionamos el servidor de España (Spain). Se selecciona **tm** y al dar "Entrar" se realizará la carga e instalación de **tm** y los paquetes auxiliares.

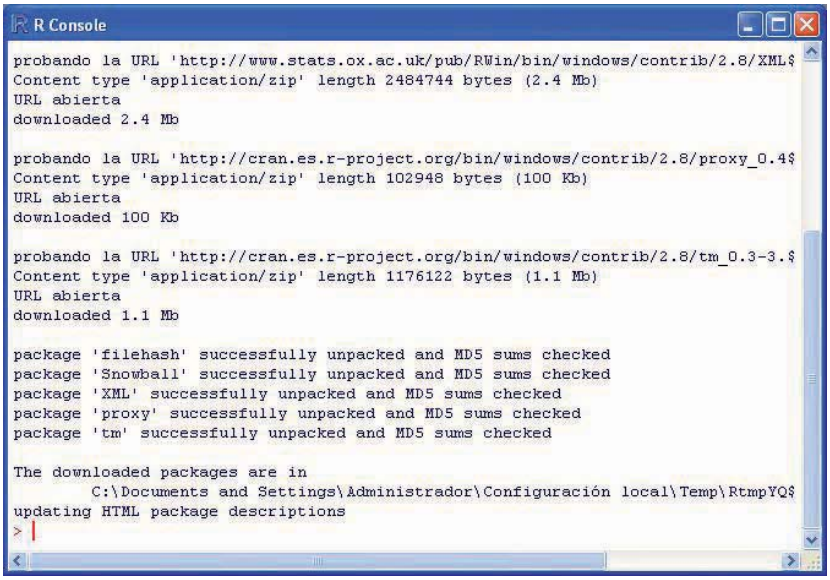


Figura 10.1. Instalación de tm y paquetes auxiliares.

3. Para confirmar la instalación de **tm** y sobre todo para usar esta librería pondremos en la línea de comandos:

```
>library(tm)
```

Ya se puede comenzar a usar la librería **tm** (el lector interesado puede consultar una información detallada de la misma en: http://rss.acs.unt.edu/Rdoc/library/tm/html/00Index.html).

Una vez cargado el paquete se realiza con el texto que se muestra a continuación el proceso de recuento de los términos presentes en el texto.

Sea un buen ejemplo para su hijo.

Ayúdele a su hijo a mejorar su actitud acerca de las matemáticas demostrándole como usted resuelve los deberes cotidianos que requieren el uso de las matemáticas. Estos deberes pueden incluir tareas como contar el dinero de un evento en beneficio del curso o del colegio, conciliar su cuenta corriente, o completar su declaración de impuestos. También puede contarle sobre las diferentes carreras que requieren el uso de las matemáticas tales como la arquitectura, la medicina, el diseño de ropa, la administración de un restaurante y la programación computacional.

Ayúdele a utilizar las matemáticas todos los días.

Motive a su hijo a resolver problemas que requieren el uso de las matemáticas en aéreas fuera del ámbito escolar. En el supermercado, pídale que calcule cuánto cuestan los cuatro tarros de atún u otros comestibles. En el automóvil, pregúntele cuanto tiempo se demoraran para llegar a su destino basándose en la velocidad. En la tienda de juguetes, pídale que calcule el precio de algún juguete que este con descuento y cuanto dinero necesitaría ahorrar durante cuánto tiempo hasta que pueda comprarlo.

Aprenda cuales son las normas académicas requeridas.

Es importante saber cuáles son las habilidades que su hijo debe aprender para el curso en que se encuentra. Puede averiguar las normas académicas para el grado de su hijo a través del sitio web del Ministerio de Educación o pedírselos al profesor(a) de su hijo. Si usted está al tanto de lo que su hijo estará aprendiendo, le será más fácil reforzar estas habilidades con las actividades apropiadas en su casa.

Manténgase al tanto del tipo de tarea de matemáticas que le dan.

La tarea de matemáticas de su hijo si contiene problemas repetitivos o incluye problemas que requieren más creatividad, como Los Problemas de la Semana

del Calendario Matemático (5 a 8 Básico), que ponen a prueba el nivel de comprensión de su hijo de los conceptos matemáticos. Pregúntale al profesor(a) que técnicas utiliza para ayudar a los estudiantes a sentirse más hábiles en las matemáticas.

Ponle atención los detalles.

Usted puede ayudar a su hijo con su tarea de matemáticas enseñándole que para resolver los ejercicios y/o problemas debería notar cada paso en la operación como también los cálculos matemáticos además de la respuesta. Su hijo debería revisar su tarea para estar seguro de que no exista ningún error. Este detalle le ayudaría a ver claramente donde estuvo su error. Es una buena idea si pone un límite a las distracciones y fija la misma hora cada día para hacer las tareas.

Disfrute de juegos de matemáticas en casa.

Existen muchos juegos que usted puede disfrutar con su hijo que involucran el uso de las matemáticas. Desde los cursos básicos, su hijo puede aprender a disfrutar de las matemáticas a través de juegos como el ajedrez, domino, naipes, damas, dados, etc.

Lea libros que incluyen las matemáticas.

La Reforma Educacional integra las diversas materias en el currículo escolar para que los estudiantes puedan aprender mejor como estas materias se relacionan entre sí. Como puede uno incluir las matemáticas en un curso de Historia o Ingles. Una buena manera es el leer libros cuyos personajes principales resuelvan problemas utilizando las matemáticas o la lógica.

Para realizar dicho proceso se crean 7 ficheros de texto cada uno de los cuales contiene una frase que comienza en el texto señalado con negrita.

El fichero de sintaxis para realizar el análisis es el siguiente:

```
setwd(“E:/curso 2009-10/tecnicas e instrumen/análisis contenido”)
library(tm) #carga la librería tm
```

```
vignette("tm") # carga una "viñeta", extracto en pdf de las
funciones de tm
txt<-DirSource(directory = "prueba",encoding = "LATIN1")

(docu <- Corpus(txt,readerControl = list(reader = read-
Plain,language = 'spanish'))) # se almacena el texto en la
variable docu, si hubieran sido varios textos corpus los agru-
pa en uno

summary(docu) # resume las característica del documento
inspect(docu[1:7]) # se inspecciona los textos

tdm<-TermDocumentMatrix(docu,control = list(stopwords = TRUE))
# términos de todos los documentos

inspect(tdm[1:20]) # inspecciona los 20 primeros términos
findFreqTerms(tdm,3) # localiza los términos que se repiten 3
veces

dtm <- DocumentTermMatrix(docu, control = list(stopwords =
TRUE)) # términos por documento

inspect(dtm[1:7])


frec<-findFreqTerms(dtm,3) # localiza los términos que se repi-
ten 3 veces

findAssocs(dtm, "hijo", 0.57) # muestra los términos cuyo gra-
do de asociación con el término "hijo" es mayor o igual a 0,57

inspect(removeSparseTerms(dtm, 0.4))

x<-as.matrix(tdm)# almacena los términos y su frecuencia en una
matriz
write.csv(x,"docu1.csv") # salva el fichero de terminos con su
frecuencia para un posterior uso
Datos1 <-
  read.table("E:/curso 2009-10/tecnicas e instrumen/análisis
contenido/docu1.csv",
  header=TRUE,  sep=",",  na.strings="NA",  dec=".",
strip.white=TRUE)# importa el fichero de datos

HClust.1 <- hclust(dist(model.matrix(~-1 +
```

```
prue1.txt+prue2.txt+prue3.txt+prue4.txt+prue5.txt+prue6.txt+pru
e7.txt, Datos1)) , method= "ward")# realiza un análisis de
cluster

plot(HClust.1, main= "Cluster Dendrogram for Solution
HClust.1", xlab="Observation Number in Data Set Datos1",
sub="Method=ward;
  Distance=euclidian")

summary(as.factor(cutree(HClust.1, k = 2))) # Cluster Sizes
by(model.matrix(~-1 + prue1.txt + prue2.txt + prue3.txt +
prue4.txt + prue5.txt + prue6.txt + prue7.txt, Datos1),
as.factor(cutree(HClust.1, k = 2)), mean) # Cluster Centroides

biplot(princomp(model.matrix(~-1 + prue1.txt + prue2.txt +
prue3.txt + prue4.txt + prue5.txt + prue6.txt + prue7.txt,
Datos1)), xlabs =
  as.character(cutree(HClust.1, k = 2)))

save("Datos1",file="E:/curso 2009-10/tecnicas e instru-
men/análisis contenido/Datos1.rda")

Datos1$X[as.factor(cutree(HClust.1, k = 2))==1] # términos de
cada cluster
Datos1$X[as.factor(cutree(HClust.1, k = 2))==2]
```

Los resultados más relevantes son los siguientes:

> inspect(tdm[1:20]) # **inspecciona los 20 primeros términos**
A term-document matrix (20 terms, 7 documents)

Non-/sparse entries: 23/117
Sparsity : 84%
Maximal term length: 14
Weighting : term frequency (tf)

```
                Docs
Terms            prue1.txt prue2.txt prue3.txt prue4.txt prue5.txt prue6.txt prue7.txt
  académicas             0         0         2         0         0         0         0
  acerca                 1         0         0         0         0         0         0
  actitud                1         0         0         0         0         0         0
  actividades            0         0         1         0         0         0         0
  además                 0         0         0         0         1         0         0
  administración         1         0         0         0         0         0         0
  aéreas                 0         1         0         0         0         0         0
  ahorrar                0         1         0         0         0         0         0
  ajedrez,               0         0         0         0         0         1         0
  ámbito                 0         1         0         0         0         0         0
  aprenda                0         0         1         0         0         0         0
  aprender               0         0         1         0         0         1         1
  aprendiendo,           0         0         1         0         0         0         0
  apropiadas             0         0         1         0         0         0         0
  arquitectura,          1         0         0         0         0         0         0
  atención               0         0         0         0         1         0         0
  atún                   0         1         0         0         0         0         0
  automóvil,             0         1         0         0         0         0         0
  averiguar              0         0         1         0         0         0         0
  ayudar                 0         0         0         1         1         0         0
```

> findFreqTerms(tdm,3) # **localiza los términos que se repiten 3 veces**
[1] "hijo" "problemas" "juegos"

> findAssocs(dtm, "hijo", 0.57) # **muestra los términos cuyo grado de asociación con "hijo"es mayor o igual a 0,57**

```
      hijo        usted  académicas  actividades     aprenda  aprendiendo,  apropiadas
       1.0          0.7         0.6          0.6         0.6           0.6         0.6
 averiguar       cuales      cuáles         debe   educación    encuentra.       fácil
       0.0          0.6         0.6          0.6         0.6           0.6         0.6
     grado  habilidades       hijo.   importante  ministerio        normas  pedírselos
       0.0          0.6         0.6          0.6         0.6           0.6         0.6
  reforzar  requeridas.       sitio          web
        06          0.6         0.6          0.6
```

Como se puede observar por los resultados anteriores los términos que aparecen con mayor frecuencia son: hijo, problemas y juegos. Por tanto, está claro que hay dos formas de fomentar el gusto por las matemáticas en nuestros hijos mediante la realización de problemas y mediante el juego.

Para continuar con los análisis nos planteamos cual sería la proximidad de los 7 párrafos en función de la presencia o ausencia de cada uno de los

197 términos que tuvo el corpus o conjunto de los 7 párrafos. Para realizar tal proceso se utilizó el análisis de cluster o agrupamiento (información de ampliación para el estudiante).

```
> HClust.1 <- hclust(dist(model.matrix(~-1 +
+ prue1.txt+prue2.txt+prue3.txt+prue4.txt+prue5.txt+prue6.txt+prue7.txt,
+ Datos1)) , method= "ward")
```

```
> plot(HClust.1, main= "Cluster Dendrogram for Solution HClust.1", xlab=
+ "Observation Number in Data Set Datos1", sub="Method=ward;
+ Distance=euclidian")
```

```
> summary(as.factor(cutree(HClust.1, k = 2))) # Cluster Sizes
  1  2
165  32
```

Según se observa se forman dos cluster, el primero, el mayoritario, que contiene 165 términos y el segundo con tan sólo 32 términos. Como el gráfico Dendrograma resultaba poco esclarecedor se realizó un biplot.

```
> by(model.matrix(~-1 + prue1.txt + prue2.txt + prue3.txt + prue4.txt +
+ prue5.txt + prue6.txt + prue7.txt, Datos1), as.factor(cutree(HClust.1, k =
+ 2)), mean) # Cluster Centroids
```

```
INDICES: 1
prue1.txt prue2.txt prue3.txt prue4.txt prue5.txt prue6.txt prue7.txt
0.27272727 0.04848485 0.21212121 0.21818182 0.26666667 0.15151515
0.21212121
_____________________________________________________________
INDICES: 2
prue1.txt prue2.txt prue3.txt prue4.txt prue5.txt prue6.txt prue7.txt
   0.000     1.125     0.000     0.000     0.000     0.000     0.000
```

Los centros de cada texto en los 2 cluster vienen representados por el recuadro anterior donde se puede observar como todos los textos excepto el párrafo 2 se agrupan en el cluster 1 y el texto 2 se va al cluster 2.

```
> biplot(princomp(model.matrix(~-1 + prue1.txt + prue2.txt + prue3.txt +
+ prue4.txt + prue5.txt + prue6.txt + prue7.txt, Datos1)), xlabs =
+ as.character(cutree(HClust.1, k = 2)))
```

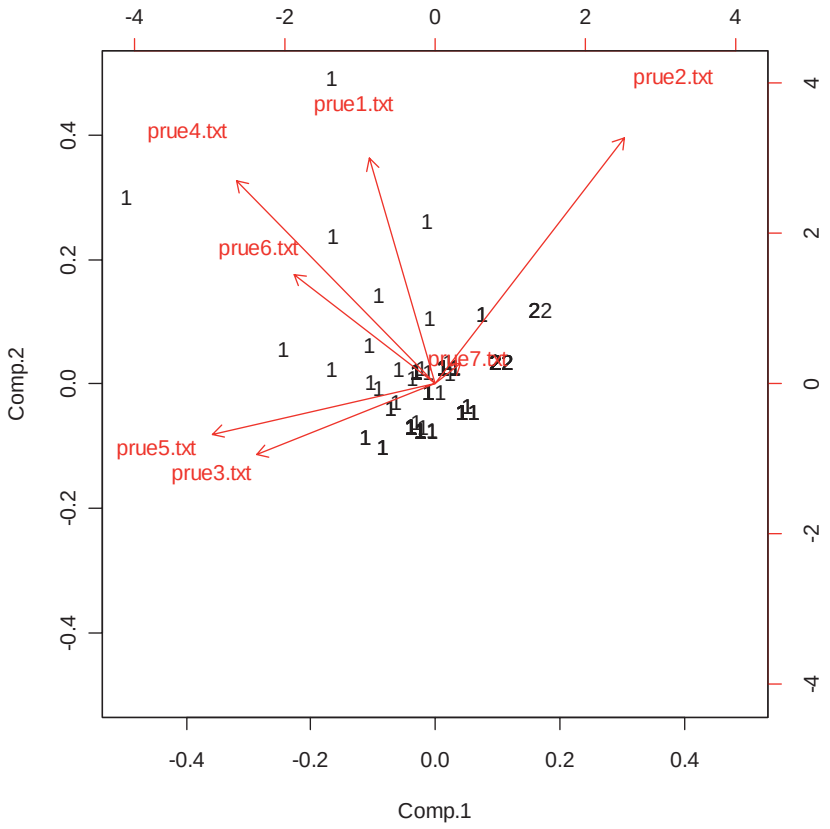


Figura 10.2. Gráfico vectorial de los párrafos en los cluster.

Los términos del cluster 1 son los siguientes:

```
> Datos1$X[as.factor(cutree(HClust.1, k = 2))==1]
[1] académicas  acerca  actitud  actividades  además
[6] administración  ajedrez,  aprenda  aprender  aprendiendo,
[11] apropiadas  arquitectura,  atención  averiguar  ayudar
[16] ayudaría  ayúdele  básico),  básicos,  beneficio
[21] buen  buena  cálculos  calendario  carreras
[26] casa.  claramente  colegio,  completar  comprensión
[31] computacional.  conceptos  conciliar  contar  contarle
[36] contiene  corriente,  cotidianos  creatividad,  cuales
[41] cuáles  cuenta  currículo  curso  cursos
[46] cuyos  dados,  damas,  dan.  debe
```

```
[51] deberes debería declaración demostrándole detalle
[56] detalles. día diferentes dinero diseño
[61] disfrutar disfrute distracciones diversas domino,
[66] educación educacional ejemplo ejercicios encuentra.
[71] enseñándole error. escolar estudiantes etc.
[76] evento exista existen fácil fija
[81] grado hábiles habilidades hijo hijo.
[86] historia hora idea importante impuestos.
[91] incluir incluye incluyen ingles. integra
[96] involucran juegos lea leer libros
[101] límite lógica. manera manténgase matemáticas
[106] matemáticas. matemático matemáticos matemáticos. materias
[111] medicina, mejor mejorar ministerio misma
[116] naipes, ningún nivel normas notar
[121] operación paso pedírselos personajes pone
[126] ponen ponle pregúntale principales problemas
[131] profesor(a) programación prueba puedan reforma
[136] reforzar relacionan repetitivos requeridas. requieren
[141] resolver respuesta. restaurante resuelvan resuelve
[146] revisar ropa, seguro semana sentirse
[151] sí. sitio tales tarea tareas
[156] tareas. técnicas tipo través usted
[161] utiliza utilizando ver web y/o
197 Levels: académicas acerca actitud actividades además ... y/o
```

Los términos del cluster 2 son los siguientes:

```
> Datos1$X[as.factor(cutree(HClust.1, k = 2))==2]
[1] aéreas ahorrar ámbito atún automóvil,
[6] basándose calcule comestibles. comprarlo. cuanto
[11] cuánto cuatro cuestan demoraran descuento
[16] destino días. escolar. juguete juguetes,
[21] llegar motive necesitaría pídale precio
[26] pregúntele pueda supermercado, tarros tienda
[31] utilizar velocidad.
197 Levels: académicas acerca actitud actividades además ... y/o
```

BIBLIOGRAFÍA

Bardin, L. (1977). *Análisis de contenido.* Madrid: Akal.

Berelson, B. (1952). *Content Analysis in Comunication Research*. Free Press, Glencoe.

Bosch, J. y Lluís C. (1993). *Encuestas telefónicas y por correo*. Madrid: Centro de Investigaciones Sociológicas.

Bruner, Jerome (1991). *Actos de significado. Más allá de la revolución cognitiva.* Madrid: Alianza.

Canales, M. y Peinado, A. (1994). Grupos de Discusión En J. M. Delgado y J. Gutiérrez (coor.) *Métodos y Técnicas cualitativas de investigación en Ciencias Sociales.* Madrid: Sintesis.

Cohen, L. y Manion, L. (1990). *Métodos de investigación educativa.* Madrid: La Muralla.

Colle, R. (1988). *ANATEX: Software de análisis de concurrencias.* Esc. De Periodismo PUC y Secico. Santiago de Chile.

Duverger, M. (1974). *Sociología de la Política.* Elementos de Ciencia Política Barcelona: Ariel.

Flick, Uwe (2004). *Introducción a la investigación cualitativa.* Madrid: Morata.

Fontana, A. y J. H. Frey (1994). Interviewing. The Art of Science, en N. K. Denzin y Y. S. Lincoln (eds.) *Handbook of Qualitative Research*. London: SAGE.

Groves, R. M. (1989). "Survey errors and survey costs". New York, Wiley.

Hyman, Herbert (1984). *Diseño y análisis de las encuestas sociales*. Amorrortu. Buenos Aires.

Ibáñez, J. (1979). *Más allá de la Sociología.* El grupo de discusión: Técnica y Crítica. Madrid: Siglo XXI.

Igúzquiza, Vidal (2002). *Tipos de encuestas y diseños de investigación.* Universidad Pública de Navarra, D.L. Pamplona.

Kish, Leslie (1995). *Diseño estadístico para la investigación*. Siglo XXI. Madrid.

Krippendorff, K. (1990). *Metodología del análisis de contenido. Teoría y práctica.* Paidós comunicación.

Krueger, R. A. (1988). *El grupo de discusión. Guía práctica para la investigación aplicada.* Madrid: Pirámide.

McCracken, G. (1988). *The Long Interview (Qualitative Research Methods).* London: Sage.

Patton, H. (1980). *Qualitative Evaluation.* London: Sage.

Spradley, J. P. (1979). *The Ethnographic Interview.* New York: Holt, Rinehart and Winston.

Stubbs, M. (1987). *Análisis del discurso. Análisis socio-lingüístico del lenguaje natural.* Madrid: Alianza Editorial.

Taylor, S. J. y Bogdan, R. (1994). *Introducción a los métodos cualitativos de investigación.* Barcelona: Paidós.

Tuckman, B. W. (1972). *Conducting Educational Research*. New York: Harcourt Brace Jovanovich.

Vázquez Recio, R. y Angulo Rasco, F. (2003). *Introducción a los estudios de casos. Los primeros contactos con la investigación etnográfica*. Archidona: Aljibe.

Walter, Rob (1989). *Métodos de investigación para el profesorado*. Madrid: Morata.

Woods, P. (1987). *La escuela por dentro. La etnografía en la investigación educativa.* Barcelona/Madrid: Paidós/MEC.

Whyte, W. F. (1982). *Learning from the field: a guide from experience.* Beverly Hills: Sage.

Al realizar cualquier investigación en Educación, ya sea en el contexto del aula o en cualquier otro, se necesita utilizar instrumentos y emplear técnicas acordes con los mismos. Son varios los condicionantes para utilizar uno u otro instrumento y usar una u otra técnica. El profesional de la Educación tiene que saber las ventajas e inconvenientes de cada técnica e instrumento, debe conocer las limitaciones, necesita calcular y dar a conocer la calidad técnica de los instrumentos que utiliza, necesita justificar sus bondades para añadir a su trabajo profesional el rigor científico.

El texto aborda las distintas técnicas e instrumentos para recoger la información. Comienza justificando la medida en educación, aborda después los condicionantes técnicos de los instrumentos y luego repasa cada una de las principales técnicas e instrumentos de medida: la observación, test y pruebas objetivas, la encuesta, la entrevista, las escalas, las pruebas criteriales, las técnicas grupales de recogida de información y finalmente el análisis de documentos. No se ha pretendido profundizar en cada uno de los tópicos presentados, ya que cada uno de ellos podría constituir un texto independiente, pero sí dar al estudiante una visión global de los mismos. La presentación de ejemplos y la referencia a investigaciones donde se utilizan las técnicas mencionadas ayudarán a la comprensión del libro.

Juan Antonio Gil Pascual, profesor titular de Métodos de Investigación en Educación en el Departamento M.I.D.E. de la Facultad de Educación de la UNED. Matemático (Estadística e Investigación Operativa) y doctor en Educación. Con dilatada docencia en todos los niveles educativos y como experto en estudios de opinión en el Área de Recursos Humanos. Especialista en análisis y tratamiento de datos de carácter multivariante con paquetes estadísticos (SPSS,R). Autor de publicaciones y ponencias de aplicaciones de metodologías estadísticas al ámbito social. Colaborador en numerosos proyectos de investigación como analista de datos. Premio Nacional de Investigación Educativa (1998).

UNED Editorial

colección Grado

ISBN: 978-84-362-6250-6
02205
9 788436 262506

6302205GR01A01